

Promoters: Prof. Dr. ir. Wim M. Cornelis
UNESCO Chair of Eremology
Department of Soil Management
Faculty of Bio-Science Engineering, Ghent University

Prof. Dr. Eric Van Ranst
Department of Geology and Soil Science
Faculty of Sciences, Ghent University

Prof. Dr. ir. Paul Mafuka Mbe-Mpie
Department of Natural Resources Management
Faculty of Agronomy, University of Kinshasa

Dean: Prof. Dr. ir. Guido Van Huylenbroeck

Rector: Prof. Dr. Paul Van Cauwenberge

ir. Yves-Dady BOTULA MANYALA, MSc.

Indirect methods to predict hydrophysical properties of soils of Lower Congo

Thesis submitted in fulfilment of the requirements for the degree of Doctor (PhD) in
Applied Biological Sciences: Land and Forest Management

Illustration on the cover: Photo of a deep roadcut near the bridge of the Kwilu river in Bas-Congo (15th December 2009).

Citation:

Botula, Y.-D. 2013. Indirect methods to predict hydrophysical properties of soils of Lower Congo. PhD thesis, Ghent University.

ISBN-number: **978-90-5989-648-2**

The author and the promoters give the authorisation to consult and to copy parts of this work for personal use only. Every other use is subject to the copyright laws. Permission to reproduce any material contained in this work should be obtained from the author.

To the memory of my father,
Dr. ir. Michel Botula Manyala ma Bopoto

Acknowledgements

My doctoral studies would have not be completed without the assistance of many people.

I am particularly grateful to Ghent University for providing the Special Research Fund (BOF) scholarship to support my research under the contract BOF08/DOS/051.

I express my sincere gratitude to my first promoter, Prof. Dr. ir. Wim M. Cornelis for his comprehensive guidance, constructive criticisms and great support throughout my doctoral studies. Prof. Wim, I thank you for giving me the opportunity to carry out this research on pedotransfer functions to predict water retention of soils of the humid tropics. Your great scientific knowledge, your kindness and positive attitude have been a great source of inspiration and encouragement. I really enjoyed working under your supervision and I have benefited a lot from our collaboration. Thank you for having the door of your office open for me for all kinds of questions or problems!

I am deeply grateful to my promoter, Prof. Dr. Eric Van Ranst for his constructive criticisms and valuable suggestions which greatly improved the quality of this work. Prof. Van Ranst, I thank you for your support during my studies at Ghent University. I have benefited a lot from your excellent teaching skills, your great knowledge and experience about soils of the humid tropics and your readiness to help.

Many thanks to my promoter and local supervisor Prof. Dr. ir. Paul Mafuka Mbe-Mpie for useful advices during my research and for his personal involvement and support during the fieldwork campaigns in Bas-Congo and in Kinshasa. Prof. Mafuka, I thank for your help and for the interesting discussions we had around my PhD work.

The constructive comments and suggestions of the members of the reading committee: Dr. ir. Henk Wösten, Prof. Dr. ir. Ann Verdoodt and Dr. ir. Jan Baetens that greatly improved the quality of the doctoral dissertation are deeply appreciated. Your valuable inputs in various aspects of this PhD work are also greatly acknowledged.

My special thanks to Prof. Dr. ir. Geert Baert whose expertise on soils of Lower Congo and deep knowledge of the Bas-Congo region have been of great help particularly during the fieldwork campaigns. His valuable inputs at different stages of this work are greatly appreciated.

I would like particularly to thank Dr. Attila Nemes who played a key role in the success of my PhD work. His valuable inputs and suggestions during my stay in Beltsville (Maryland) and beyond have greatly improved the scientific quality of this thesis. His readiness to help and positive attitude gave me the boost I needed. I extend my gratitude to his wife Aniko for her kindness and her hospitality. Thank you Attila and thank you Aniko for the wonderful and unforgettable moments I had with you!

I would like specially to thank all the current members of the Soil Physics Unit: Maarten De Boever, Jasmien Wildemeersch, Jan De Pue, Nguyen Minh Phuong, Tran Ba Linh, Muhammed Khlosi, Bashar Al-Barri, Thadeo Mpulila, Geoffrey Waweru Nyambura, Medrina Linda Kamwendo, Mansonia Alejandra Pulido Moncada, Mojtaba Pakparvar, Meisam Rezaei, Hari Ram Upadhayay, for their

support in various ways and the nice atmosphere at work and during the break. I enjoyed your company guys! I also express my gratitude to Jan Vermang for his kindness, his hospitality and his precious help.

I would like to extend my gratitude to my colleagues in other research units of our department particularly Philip De Smedt, Islam Monirul, Jeroen De Waele, Mesfin Tsegaye Gebremikael, Kanagaratnam Jegajeevagan, Nele Delbecque and Nick Ryken for their kindness, their support and the nice moments we shared.

Un très grand merci au personnel de l'Institut National pour l'Etude et la Recherche Agronomique (INERA) à Luki, Gimbi et M'Vuazi, de la Compagnie Sucrière de Kwilu-Ngongo et de la Station Phytotechnique de la Faculté d'Agronomie de l'UNIKIN pour leur aide précieuse pendant la récolte de mes échantillons de sol tant dans le Bas-Congo qu'à Kinshasa.

I would like to thank the lab technicians: Maarten Volckaert, Thérèse Buyens, Anne-Mie Terryn, Luc Deboosere from the Soil Management Department and Nicole Vindevogel and Veerle Vandenhende from the Laboratory of Soil Science for their precious help in the different analyses of my soil samples.

Je voudrais remercier mes petites sœurs Tania, Christelle, Margot, Prisca et mon petit frère Hénoc pour leur soutien, leur patience et leurs prières pendant toutes ces années. Je voudrais dire également un grand merci à tous mes membres de famille tant du côté paternel que du côté maternel qui, de près ou de loin et en diverses occasions, m'ont témoigné de leur amour et de leur soutien. Ne préférant pas vous nommer de peur d'en oublier certains, prière de trouver dans ces quelques mots l'expression de ma profonde reconnaissance.

Du fond de mon coeur, je voudrais témoigner ma profonde gratitude à mes parents pour leur amour, leurs prières et leur soutien pendant toutes ces années. Je salue la mémoire de mon père Michel Botula Manyala qui m'a donné goût à la science et à la recherche. Son départ soudain moins de cinq mois après le début de mes études doctorales fut ma plus grande douleur. J'aurai tant voulu qu'il en voie le dénouement. Je dédie cette thèse à sa mémoire.

A toi ma mère, Sabine Ngandu Kapinga, je dis infiniment merci pour toutes ces années de sacrifices et pour l'amour et le soutien inconditionnels que tu n'as cessé de me témoigner.

Finally, I would like to thank all my friends and all the persons who have been a blessing for me in various ways and under various circumstances throughout my life. You are too many to name but please may you find in these few words the expression of my deep gratitude.

May God bless you all!

Yves-Dady Botula Manyala,
Ghent, September 2013

Table of contents

Summary.....	vii
List of figures.....	xi
List of tables.....	xv
List of symbols.....	xix
List of abbreviations.....	xxi
1 General introduction.....	1
1.1 Background and nature of the problem.....	3
1.2 Knowledge gaps	5
1.3 Objectives of the study	7
1.4 Outline of the dissertation.....	8
2 Description of the study area: Lower Congo	11
2.1 Biophysical setting	13
2.1.1 Geographical location	13
2.1.2 Climate.....	14
2.1.3 Geology.....	14
2.1.4 Topography	15
2.1.5 Vegetation	15
2.2 Soils	15
2.2.1 Previous pedological studies in Lower Congo	15
2.2.2 Systematic soil survey in Lower Congo.....	16
2.2.3 Soil samples used for pedotransfer modelling.....	16
2.2.4 Soil samples used for mapping soil hydrophysical properties in Lower Congo.....	19
2.3 Socio-economic setting.....	19
3 Pedotransfer functions to predict water retention of soils of the humid tropics: a review	21
3.1 Introduction.....	23
3.2 Categorization of water retention PTFs	24
3.2.1 Class and continuous PTFs.....	24
3.2.2 Point-based PTFs, parameter-based PTFs and ‘pseudo-continuous’ PTFs.....	25
3.2.3 Specific approach-based PTFs.....	28
3.2.4 Equation-based PTFs and pattern-recognition PTFs.....	36

3.3	Summary.....	38
4	Particle size distribution models for soils of the humid tropics	47
4.1	Introduction.....	49
4.2	Materials and methods	51
4.2.1	Soil datasets.....	51
4.2.2	Particle size distribution models	52
4.2.3	Fitting procedure.....	53
4.2.4	Criteria for model comparison.....	53
4.3	Results and discussion	55
4.3.1	Estimates of parameters for various PSD models	55
4.3.2	Comparison of the fitting ability of the PSD models	58
4.3.3	Influence of texture on the fitting ability of the PSD models	62
4.3.4	Evaluation of prediction ability of the best fitting PSD models	66
4.4	Conclusions.....	71
5	Evaluation of pedotransfer functions for predicting water retention of soils in Lower Congo (D.R. Congo).....	72
5.1	Introduction.....	74
5.2	Materials and methods	76
5.2.1	The study area.....	76
5.2.2	Evaluation dataset	76
5.2.3	Pedotransfer functions considered in this study	79
5.2.4	Routine soil analyses.....	82
5.2.5	Soil water retention	82
5.2.6	Evaluation methods.....	82
5.3	Results and discussion	83
5.3.1	Exploratory study on potential predictors of water retention parameters	83
5.3.2	Evaluation of the selected pedotransfer functions	88
5.4	Conclusions.....	93
6	Prediction of water retention of soils from the humid tropics by the nonparametric k-nearest neighbour approach.....	95
6.1	Introduction.....	97
6.2	Materials and methods	100
6.2.1	Soil datasets.....	100
6.2.2	The k-nearest neighbour technique	102
6.2.3	Design parameters k and p for the k-nearest neighbour algorithm	104

6.2.4	Ensemble of k-nearest neighbour estimations	104
6.2.5	Input and output attributes used.....	105
6.2.6	Evaluation criteria.....	105
6.2.7	Comparison with two published ‘tropical’ pedotransfer functions.....	106
6.3	Results and discussion	107
6.3.1	Ensembles of k-nearest neighbour estimations.....	107
6.3.2	Optimizing the <i>k</i> and <i>p</i> terms.....	109
6.3.3	Prediction of water retention from an international ‘tropical’ database.....	112
6.3.4	Prediction performance of the k-nearest neighbour and multiple linear regression approaches	119
6.4	Conclusions.....	122
7	Hierarchical pedotransfer functions to predict bulk density of Lower Congo soils	123
7.1	Introduction.....	125
7.2	Materials and Methods	127
7.2.1	Study area: Lower Congo.....	127
7.2.2	Lower Congo soil dataset.....	127
7.2.3	Predictive capability of published bulk density pedotransfer functions on the Lower Congo dataset.....	129
7.2.4	Development of pedotransfer functions to predict bulk density	129
7.3	Results and discussion	132
7.3.1	Predictive ability of published BD PTFs applied on the Lower Congo dataset	132
7.3.2	Hierarchical multiple linear regression PTFs to predict bulk density	135
7.3.3	Pedotransfer functions based on pattern-recognition to predict bulk density.....	140
7.4	Conclusion	143
8	Hierarchical pedotransfer functions to predict water retention of soils in Lower Congo	145
8.1	Introduction.....	147
8.2	Materials and Methods	148
8.2.1	Study area: Lower Congo.....	148
8.2.2	The Lower Congo soil dataset	148
8.2.3	Development of PTFs to predict water retention.....	149
8.3	Results and discussion	151
8.3.1	Hierarchical MLR PTFs to predict water retention	151
8.3.2	PTFs based on pattern-recognition to predict water retention.....	158
8.3.3	Comparison between equation-based approach and pattern-recognition approach in prediction of water retention	159

8.4	Conclusion	162
9	Performance of three pore-solid fractal approaches for soil water retention of highly weathered soils of the humid tropics	163
9.1	Introduction.....	165
9.2	Materials and Methods	167
9.2.1	Study area: Lower Congo.....	167
9.2.2	Soil datasets.....	167
9.2.3	Grouping of soils per textural classes	168
9.2.4	Pore-solid fractal models used for fitting and prediction of soil water retention function 169	
9.3	Results and discussion	173
9.3.1	Relationships between the fractal dimensions of PSD and SWRC.....	173
9.3.2	Fitting performance of three pore-solid fractal models tested on 196 soil samples of the Lower Congo dataset	177
9.3.3	Prediction performance of the symmetric PSF model calibrated on the IGBP-Trop-2 dataset and tested on Lower Congo soils	178
9.3.4	Prediction performance of the symmetric PSF model tested on Lower Congo soils using k-fold cross validation	181
9.4	Conclusion	184
10	Evaluation of pedotransfer functions for mapping soil hydrophysical properties in Lower Congo 185	
10.1	Introduction.....	187
10.2	Materials and Methods	188
10.2.1	Study area: Lower Congo.....	188
10.2.2	Datasets	189
10.2.3	Site-specific pedotransfer modelling approaches	190
10.2.4	Published pedotransfer functions for water retention	191
10.2.5	Calculation of depth-weighted averages of clay, sand, OC, BD, and SWHC	191
10.2.6	Use of existing Lower Congo soil maps	192
10.2.7	Mapping of physico-chemical and hydrophysical properties	192
10.3	Results and discussion	195
10.3.1	Preliminary statistical exploration of the target soil properties	195
10.3.2	Spatial variability of Clay _{avg} , Silt _{avg} , Sand _{avg} , pH _{avg} and OC _{avg}	196
10.3.3	Comparison between ordinary kriging and inverse distance weighting.....	198
10.3.4	Effects of the empirical PTFs-estimated values on the quality of inverse distance weighting and kriging interpolations	200

10.3.5	Mapping soil bulk density in Lower Congo	201
10.3.6	Mapping PSWHC in Lower Congo	203
10.4	Conclusion	205
11	General discussion, conclusions and future research.....	207
11.1	General discussion.....	208
11.2	Conclusion	212
11.3	Future research	214
References		215
Curriculum vitae.....		237

Summary

Soil and water constitute two important natural resources. They are fundamental supports of natural and artificial ecosystems and human life. Soil and water are intimately related in nature. Therefore, a deep understanding of their properties, behaviours and interactions is essential for their rational management. The emerging interdisciplinary research field of hydropedology attracts substantial attention because of its promise to bridging pedology and hydrology. Soil hydraulic properties, such as the soil water retention curve (SWRC) and the hydraulic conductivity (K), are physical properties that describe the soil-water relationship. Soil bulk density (BD) and SWRC which are the main focus of this dissertation are designated as soil hydrophysical properties.

Direct measurements of soil hydrophysical properties are still laborious, expensive and time-consuming especially for environmental studies at a regional scale. Therefore, an attractive alternative to the direct and often cumbersome measurements of soil hydrophysical properties is their estimation by so-called pedotransfer functions (PTFs). Hydraulic PTFs are predictive functions which relate more easily measurable soil data such as soil texture (sand, silt and clay), organic matter (OM) or organic carbon (OC) content and/or other data routinely measured or registered in soil surveys with hydrophysical parameters such as BD and the SWRC. During the last decades, many PTFs have been developed to predict BD and SWRC for soils of temperate regions and only few PTFs have been developed for soils of the humid tropics. There is an urgent need to provide reliable information on these hydrophysical properties through their estimation by PTFs. To date, no hydrophysical PTFs have been developed for soils from the D.R. Congo in general and for soils in Lower Congo in particular.

The overall objective of this research was to test and develop indirect methods (PTFs) for predicting hydrophysical properties of soils in Lower Congo region.

In Chapter 2, the study area was briefly described. The term ‘Lower Congo’ is used to indicate the western part of the D.R. Congo that is located between the Congo River and the Atlantic Ocean. Lower Congo lies between the latitudes 4° and 5° S and the longitudes 13° and 16° E, covering a total area of almost 64,000 km². Intensive soil survey campaigns have been carried out since the 1940s. During those surveys, more than 300 soil profiles were described for this region with more than 1,400 soil samples being analysed for various physico-chemical properties at the Laboratory of Soil Science at Ghent University (Belgium).

In Chapter 3, an extensive review of existing PTFs to predict water retention of soils of the humid tropics was made based on available literature. Most of the local ‘tropical’ PTFs published in peer-reviewed journals originated from three countries: Brazil, Nigeria and India. In other regions of the humid tropics like in Central Africa, there were no (recorded) efforts in the development of local water retention PTFs. Regarding predictor variables used, particle size distribution (PSD) or texture, OC/OM and BD were found to be the most common in tropical PTFs, whereas other soil properties that are found in various databases of soils from the humid tropics such as pH, CEC, DCB-Fe and DCB-Al, were considered to a limited extent only.

In Chapter 4, when evaluating the performance of two-, three- and four-parameter models for describing cumulative PSD of soils from the humid tropics based on a large dataset from Lower Congo, we found that several models like the two and three-parameter models of Haverkamp and Parlange (1986) (VG_2p and VG_3p), the three-parameter model of Skaggs et al. (2001) (Sk_3p) and the four-parameter equation of Gompertz tested in Nemes et al. (1999) (G_4p) were not suitable to describe PSD of soils in the humid tropics. On the other hand, the two and three-parameter models of Fredlund et al. (2000) (F_3p and F_4p), the three-parameter model of Weibull used by Assouline et al.

(1998) (W_3p) and the four-parameter function of Andersson's reported by Jauhiainen (2004) (A_4p) showed outstanding fitting performance. Based on these results, F_3p, F_4p, W_3p and A_4p models are highly recommended in order to get a better description of the PSD of soils of the humid tropics and achieve standardisation of PSD in international 'tropical' databases.

In Chapter 5, before engaging in developing specific water retention PTFs for Lower Congo soils, the ability of some 'point-based' and 'parameter-based' PTFs from various geographical origins to predict water content at -33 kPa and -1500 kPa was statistically assessed based on a limited soil dataset (60 samples) from Lower Congo. This further allowed to identify potential predictors of water retention of highly weathered soils. A number of published and well-known PTFs derived from soils of temperate-climate regions ('temperate' PTFs) and from soils of the sub(humid) tropics ('tropical' PTFs) were selected. Based on various statistical measures, it was proved that there is no one single set of PTFs that is valid for soils of Lower Congo.

In Chapter 6, a nonparametric approach, i.e. the k-NN method was applied to predict water retention of highly weathered soils in the humid tropics. A variant of the k-NN algorithm was developed and tested for soils from Lower Congo based on an international dataset of soils of the (sub)humid tropics. The results showed that this k-NN variant was able to estimate water content at eight different matric potentials (0, -1 , -3 , -10 , -20 , -50 , -250 , and -1500 kPa), i.e., from the wet to the dry range of the SWRC, with an average RMSD $< 0.046 \text{ m}^3 \text{ m}^{-3}$ when SSC + BD or SSC + BD + CEC were selected as input variables. Particularly with the best combination of input variables (SSC + BD + CEC), the overall error in predicting water retention of 139 selected Lower Congo soils was remarkably low: 0.0360 to $0.0390 \text{ m}^3 \text{ m}^{-3}$ in the dry and very wet ranges and 0.0490 to $0.0510 \text{ m}^3 \text{ m}^{-3}$ in the intermediate range (i.e., -3 to -50 kPa) of the SWRC, as compared to values reported in literature for other PTFs. Furthermore, the k-NN variant developed in this study presents as added benefits the possibility to estimate prediction uncertainty and the flexibility to incorporate new data without the need to redevelop new equations.

In Chapter 7, a set of four 'tropical' and four 'temperate' published PTFs yielded poor prediction performance with a great tendency to underestimate BD of Lower Congo soils when tested on 196 undisturbed samples. Therefore, these published PTFs were found to be impractical for the Lower Congo region and new PTFs needed to be developed. To do so, two approaches, namely an equation-based (MLR) and a pattern-recognition approach (k-NN) were tested with different levels of information on the following soil characteristics: texture, OC, pH, CEC, DCB-Fe and DCB-Al. With the same set of predictors, the k-NN method had a better prediction performance than the MLR method with an overall prediction error (in terms of root mean square difference, RMSD) for k-NN of 0.096 Mg m^{-3} when using clay, sand and OC content, and 0.108 Mg m^{-3} with OC content, and DCB-Fe and DCB-Al, compared to the MLR approach with RMSD of 0.137 and 0.126 Mg m^{-3} , respectively. On the other hand, this study demonstrated the relative importance of DCB-Fe and DCB-Al along with other commonly used soil characteristics such as OC and texture in the improvement of BD prediction in both MLR and k-NN approaches.

In Chapter 8, similarly as for developing hierarchical BD PTFs using a Lower Congo dataset, different levels of availability of information on soil characteristics were considered to develop local soil water retention PTFs, i.e. based on Lower Congo hydrophysical dataset established within this study. The variables considered were texture, BD, OC, pH, CEC, DCB-Fe and DCB-Al and again two approaches were tested. In the first approach, PTFs in the form of MLR equations were formulated to predict water content at eight different matric potentials. In the second approach, a variant of the k-NN algorithm developed previously was used. A comparative analysis of the MLR and the k-NN

approaches showed that the k-NN method had a high prediction performance with an overall prediction error (RMSD) $< 0.035 \text{ m}^3 \text{ m}^{-3}$ compared to the MLR PTFs with $\text{RMSD} < 0.046 \text{ m}^3 \text{ m}^{-3}$. Furthermore, as expected, the locally developed Lower Congo PTFs performed better than two other published ‘tropical’ PTFs namely the parameter-based PTFs of Hodnett and Tomasella (2002) and the point-based PTFs of Minasny and Hartemink (2011). On the other hand, the relative importance of soil properties like DCB-Fe and DCB-Al in the improvement of prediction of water retention in both approaches was demonstrated along with other more commonly used properties such as BD and CEC.

In Chapter 9, a semi-physical model used to characterize soil water retention based on the pore-solid fractal (PSF) theory was applied on highly weathered soils and on coarse textured soils from the humid tropics. Previous studies related to PSF models for water retention focused on soils from temperate-climate regions. Availability of detailed PSD and SWRC data from tropical soil datasets as well as the particular hydraulic behaviour of soils of the humid tropics opened opportunities to investigate the applicability of the PSF approach to describe and predict their water retention. It was found that the symmetric PSF model of Hwang et al. (2011) can relatively well describe the water retention function ($0.010 \text{ m}^3 \text{ m}^{-3} < \text{RMSD} < 0.064 \text{ m}^3 \text{ m}^{-3}$) and performs by far better compared to the asymmetric form of the model ($0.039 \text{ m}^3 \text{ m}^{-3} < \text{RMSD} < 0.092 \text{ m}^3 \text{ m}^{-3}$). This actually contrasted with what these authors found for Korean soils, in that they found the asymmetric model to outperform the symmetric one except for coarse textured soils. However, the symmetric PSF model seemed to have, compared with our PTFs developed in Chapters 6 and 8, relatively limited prediction ability ($0.035 \text{ m}^3 \text{ m}^{-3} < \text{RMSD} < 0.087 \text{ m}^3 \text{ m}^{-3}$) particularly when the range of basic soil properties such as OM in the calibration dataset ($1.04\% \pm 1.274$) is relatively different from the validation dataset ($2.26\% \pm 2.507$). Moreover, water retention of fine textured soils with high clay content was found to be more challenging to predict than for medium and coarse textured soils due among others to the complexity of the pore geometry in clayey soils and the relative importance of adsorption forces in retaining water. More studies with other soil datasets are required in order to better appreciate the potential of the PSF theory in describing and predicting SWRC of highly weathered soils.

In Chapter 10, an evaluation of PTFs' outputs for practical application (functional evaluation) was conducted. This study assessed the ability of the aforementioned PTFs to describe the spatial variability of hydrophysical properties of highly weathered soils and some coarse textured soils of Lower Congo. First, the effects of PTF-estimated values on the prediction performance of two interpolation methods (inverse distance weighting, IDW and ordinary kriging, OK) were investigated. The results showed that OK was a better interpolation method compared to IDW. On the other hand, ‘temperate’ PTFs tested showed a strong tendency to overestimate SWHC and should not be used for mapping hydraulic properties of highly weathered soils. Second, after testing some pedotransfer modelling approaches and predictive models with different sets of predictors, the k-NN method with texture and OC as predictors of BD and the k-NN method with texture and pH as predictors for SWHC were selected as the best options for mapping hydrophysical soil properties of Lower Congo soils. Subsequently, a BD map and a SWHC map for the Lower Congo region were derived. Both maps at the scale of 1:500,000 showed considerable spatial variability of soil hydrophysical characteristics that should be taken into account in various agricultural and environmental studies in order to promote rational management of the natural resources of the Lower Congo region.

The potential user can choose which pedotransfer modelling approach to apply to estimate soil BD and water retention. The MLR approach seems to be the most straightforward method to apply in order to get a quick estimation based on physico-chemical properties of a given soil sample. However, pattern-recognition techniques such as the k-NN method offer various benefits in terms of prediction accuracy,

flexibility and estimation of uncertainty. When implemented in a user-friendly software, k-NN can become as easy to apply as a MLR equation. In the near future, such an application in Matlab[®] and Python[®] programming languages will be provided to the public upon request at the Soil Physics Research Unit of Ghent University.

Research on physical and hydraulic aspects of soils in tropical regions is still quite limited. This situation has not improved that much today. The new research field of hydropedology carries the potential of bridging hydrology and pedology to respond to the lack of data on laborious and expensively determined hydraulic soil properties in temperate as well as in tropical regions. The emergence of initiatives such as the Globally Integrated-Africa Soil Information Service (AfSIS) project resulting in the formation of a newly established comprehensive international soil database for Sub-Saharan Africa will certainly provide new opportunities for interdisciplinary soil and water-related research.

List of figures

Figure 1–1. A conceptual diagram showing the relationships of hydropedology to other related disciplines (from Lin, 2003).....	5
Figure 1–2. Schematic overview of the dissertation.	9
Figure 2–1. Lower Congo is the South-western part of the D.R. Congo located between the Congo River and the Atlantic Ocean and comprises the province of Bas-Congo and the capital city of Kinshasa.	13
Figure 2–2. Lower Congo is bordered by Angola to the South, by Cabinda and the Republic of Congo to the North, by the Atlantic Ocean to the West and by the province of Bandundu to the East.	14
Figure 3–1. Schematic representation of point, parameter-based and pseudo-continuous pedotransfer functions (PTFs) (from Haghverdi et al., 2012).	25
Figure 3–2. Schematic overview of a three-layer neural network (from Schaap and Leij, 1998).	32
Figure 3–3. (a) A regression tree to predict saturated hydraulic conductivity (Ksat) from field texture grade (S = sand, LS = loamy sand, L= loam, CL= clay loam, LiC = light clay, C= clay), bulk density (BD), and clay content. Values in nodes are the predicted $\log_{10}(\text{Ksat})$ (in mm/day), values underneath the nodes are the standard deviation of prediction. (b) A hybrid regression tree with continuous piecewise functions (from McBratney et al., 2002).	34
Figure 3–4. Graphical representation of k-nearest neighbour for finding the ‘best’ match for a target soil (from Jagtap et al., 2004).	35
Figure 3–5. Categorization of pedotransfer functions (PTFs) to predict water retention. The number in brackets represents the number of publications in which a type of PTFs has been developed to predict water retention of soils in the humid tropics.	37
Figure 3–6. Percentage of occurrence of predictors of 'tropical' pedotransfer functions in publications. PSD is particle size distribution, BD is bulk density, OC is organic carbon, OM is organic matter, CEC is cation exchange capacity, Me is moisture equivalent, DCB-Fe is dithionite-citrate-bicarbonate extractable iron, DCB-Al is dithionite-citrate-bicarbonate extractable aluminium, SS, specific surface area.	39
Figure 4–1. Textural composition [clay (0–2 μm), silt (2–50 μm), and sand (50–2000 μm)] of the Lower Congo dataset. The greyish area represents the bimodal zone of the USDA soil textural triangle in which silt content is lower than both sand and clay contents.	51
Figure 4–2. Comparative fit of ten PSD models for a bimodal (a) clay soil and (b) sandy loam soil from Lower Congo. LN_2p is Simple lognormal (2 parameters), VG_2p is Van Genuchten type (2 parameters), VG_3p is Van Genuchten type (3 parameters), F_3p is Fredlund (3 parameters), F_4p is Fredlund (4 parameters), W_3p is Weibull (3 parameters), Sk_3p is Skaggs (3 parameters), G_2p is Gompertz (2 parameters), G_4p is Gompertz (4 parameters), A_4p is Andersson (4 parameters).	58
Figure 4–3. Box plots for the R^2 -adjusted (R^2 -adj), Akaike information criteria (AIC) and Relative error (E_r) percentiles of ten PSD models for 1,412 soil samples in the Lower Congo dataset.	59
Figure 4–4. (a) Percentage of soil samples (#1,412) for which R_{sq} -adj or AIC are the best for a given model. (b) Number of soil samples for which the relative error (E_r) is less than 5% for a given model. LN_2p is Simple lognormal (2 parameters), VG_2p is Van Genuchten type (2 parameters), VG_3p is Van Genuchten type (3 parameters), F_3p is Fredlund (3 parameters), F_4p is Fredlund (4 parameters), W_3p is Weibull (3 parameters), Sk_3p is Skaggs (3 parameters), G_2p is Gompertz (2	

parameters), G_4p is Gompertz (4 parameters), A_4p is Andersson (4 parameters). (b) Number of soil samples for which the relative error (Er) is less than 5% for a given model.	61
Figure 4–5. Percentage of cases for each particle size distribution (PSD) model that yielded the smallest AIC values out of a total number of soil samples for each USDA textural class, with (a) the most represented textural classes in the Lower Congo dataset and (b) the least represented textural classes in the Lower Congo dataset. The number on top of each column indicates the number of soil samples for each soil texture. C is clay, CL is clay loam, S is sand, SCL is sandy clay loam, SL is sandy loam, L is loam, LS is loamy sand, SC is sandy clay, Si is silt, SiC is silty clay, SiCL is silty clay loam, SiL is silt loam.	63
Figure 4–6. Box plots for the Akaike information criteria (AIC) percentiles of ten PSD models for 585 fine-, 483 medium-, and 344 coarse textured soil samples in the Lower Congo dataset.	65
Figure 4–7. Box plots for the Akaike information criteria (AIC) percentiles of ten PSD models for 775 unimodal and 637 bimodal soil samples in the Lower Congo dataset.	66
Figure 4–8. Observed particle size fractions versus particle size fractions predicted by the Fredlund-type1 (F_3p), Fredlund-type2 (Fredlund_4p), Weibull (W_3p) and Andersson (A_4p) models applied on 1,412 soil samples from Lower Congo. Grey points represent unimodal soil samples (#775) and crosses represent bimodal soil samples (#637). The figure in μm above each graph represents the USDA particle size limit for which the missing fraction needs to be predicted (interpolated) when particle size fractions within other particle size limits are known.	68
Figure 4–8. Observed particle size fractions versus particle size fractions predicted by the Fredlund-type1 (F_3p), Fredlund-type2 (Fredlund_4p), Weibull (W_3p) and Andersson (A_4p) models applied on 1,412 soil samples from Lower Congo. Grey points represent unimodal soil samples (#775) and crosses represent bimodal soil samples (#637). The figure in μm above each graph represents the USDA particle size limit for which the missing fraction needs to be predicted (interpolated) when particle size fractions within other particle size limits are known (continued).	69
Figure 4–9. Difference in textural composition between the Lower Congo dataset (1,412 soil samples) used in the present study and the Korean dataset (1,385 soil samples) used by Hwang (2004).	70
Figure 4–10. Distribution of the SSE values as a function of clay content for the Fredlund-type1 (F_3p), Fredlund-type2 (Fredlund_4p), Weibull (W_3p) and Andersson (A_4p) models for 1,412 soil samples in the Lower Congo dataset.	70
Figure 5–1. Soil water retention curves of a sand, a loam, and a clay (Source: Brady and Weil, 1999).	75
Figure 5–2. Location of the sampling sites in Lower Congo.	76
Figure 5–3. Variation of clay (0–2 μm), silt (2–50 μm) and sand (50–2000 μm) content of 60 selected soil samples of the Lower Congo dataset.	77
Figure 5–4. Regression equations and coefficients of determination (r^2) describing the relation between sand, clay, DCB-Fe contents (%) with soil water content ($\text{m}^3 \text{ m}^{-3}$) of Lower Congo soils at –33 kPa and –1500 kPa.	87
Figure 5–5. Mean difference squared (MD^2) versus standard deviation of the differences squared (SDD^2) of the gravimetric pedotransfer functions. A: Arruda et al. (1987); D: Dijkerman (1988); L: Lal (1978); O: Oliveira et al. (2002); P: Pidgeon (1972).	89
Figure 5–6. Mean difference squared (MD^2) versus standard deviation of the differences squared (SDD^2) of the volumetric pedotransfer functions. AP: Aina and Periaswamy (1985); GL: Gupta and	

Larson (1979); HT: Hodnett and Tomasella (2002); RB: Rawls and Brakensiek (1982); SC: Schaap et al. (2001); VB: van den Berg et al. (1997); VE: Vereecken et al. (1989).....	90
Figure 5–7. Measured versus predicted soil gravimetric water contents at –33 kPa and –1500 kPa	91
Figure 5–8. Measured versus predicted soil volumetric water contents at –33 kPa and –1500 kPa	92
Figure 6–1. Variation of clay (0–2 μm), silt (2–50 μm) and sand (50–2000 μm) of 534 soil samples in the IGBP-Trop dataset (circles) and of 139 soil samples in the Lower Congo soil dataset (crosses) ..	102
Figure 6–2. Box plots of some physical and chemical properties of the soils of (1) IGBP-Trop (reference dataset) and (2) Lower Congo (test dataset). BD is bulk density (Mg m^{-3}), OC is organic carbon content (%) and CEC is cation exchange capacity ($\text{cmol}_c \text{ kg}^{-1} \text{ soil}$).....	107
Figure 6–3. Running root mean squared differences (RMSDs) for the Lower Congo test dataset for up to 100 ensembles using sand, silt, clay, bulk density, organic carbon, pH and cation exchange capacity as input attributes and water retention at (a) –1 kPa, (b) –20 kPa and (c) –1500 kPa as output attributes.....	109
Figure 6–4. Variations of the root mean squared differences (RMSDs) with the number of nearest neighbours k in function of p values and reference dataset sizes N_r	112
Figure 6–5. Effect of dataset size on the optimal choice of the number of selected neighbours.	112
Figure 7–1. Variation of clay (0–2 μm), silt (2–50 μm) and sand (50–2000 μm) of 196 soil samples of the Lower Congo dataset.....	128
Figure 7–2. Box plots of selected physical and chemical properties of 196 soil samples of Lower Congo. BD is bulk density (Mg m^{-3}), pH is acidity (-), OC is organic carbon content (%), CEC is cation exchange capacity ($\text{cmol}_c \text{ kg}^{-1} \text{ soil}$), DCB-Fe is dithionite-citrate-bicarbonate extractable iron (%) and DCB-Al is dithionite-citrate-bicarbonate extractable aluminium (%).	128
Figure 7–3. Measured versus predicted bulk density values of 196 soil samples from Lower Congo using four ‘tropical’ and four ‘temperate’ pedotransfer functions.....	134
Figure 7–4. Plots of residuals versus predictors for (a) PTF-1 (BD vs. clay, sand and OC) and (b) PTF-2 (BD vs. DCB-Fe, DCB-Al and OC). Sand and clay contents are in %, BD is bulk density (Mg m^{-3}), OC is organic carbon content (%), DCB-Fe is dithionite-citrate-bicarbonate extractable iron (%) and DCB-Al is dithionite-citrate-bicarbonate extractable aluminium (%).	138
Figure 7–5. Observed bulk density versus clay, sand, OC, DCB-Fe and DCB-Al contents (%) of 196 soil samples of the Lower Congo dataset	139
Figure 7–6. Measured versus predicted bulk density values of 196 soil samples from Lower Congo by the k -nearest neighbour method with eight different set of input attributes.....	142
Figure 7–7. Measured versus predicted bulk density values of 196 soil samples from Lower Congo by the multiple linear regression (MLR) and the k -nearest neighbour (k -NN) methods.	143
Figure 8–1. Measured versus predicted soil volumetric water contents at eight matric potentials of 196 soil samples from Lower Congo using the full set of available predictors in multiple linear regression PTFs.....	157
Figure 9–1. (a) Fractal dimension of particle size distribution curve versus sand content and (b) fractal dimension of soil water retention curve versus clay content for the pore-solid fractal model of Hwang et al. (2011).	174
Figure 9–2. Measured versus predicted soil water contents by the symmetric pore-solid fractal model of Hwang et al. (2011) for 196 highly weathered soils from the Lower Congo dataset used as validation dataset. Grey circles represent unimodal soils and black crosses represent bimodal soils.	181

Figure 9–3. Box plots of organic carbon content (%) of (1) IGBP-Trop-2 (calibration dataset, 209 soil samples) and (2) Lower Congo (test dataset, 196 soil samples).....	181
Figure 10–1. Location of the soil profiles and stratification of Lower Congo in two main textural zones. Crosses represent the locations of the 248 soil profiles issued from previous survey and for which only primary soil properties were available, whereas circular points represent the locations of the 24 soil profiles issued from this study and for which also hydrophysical properties have been determined. The two regions in diagonal cross represent the ‘sandy’ stratum, whereas the central region in fine points represents the ‘clayey’ stratum.	190
Figure 10–2. Histograms (bars) and theoretical normal distribution (lines) for total OC_{avg} values and log-transformed OC_{avg} values of 248 Lower Congo soil profiles.	196
Figure 10–3. Experimental isotropic semi-variograms (black circles) calculated for depth weighted average of clay ($Clay_{avg}$), sand ($Sand_{avg}$) and log-transformed depth-weighted average of silt ($\ln(Silt_{avg})$), pH ($\ln(pH_{avg})$) and OC ($\ln(OC_{avg})$) for 248 Lower Congo soil profiles. The dashed lines represent the global semi-variance.....	197
Figure 10–4. Reduction of nugget effect and the global semi-variance (dashed lines) of $Clay_{avg}$ after stratification of the study area in a ‘clayey’ zone and a ‘sandy’ zone. The grey line represents the theoretical semi-variogram (exponential model) fitted to the data located in the ‘clayey’ zone.	198
Figure 10–5. Spatial variability of bulk density ($g\ cm^{-3}$ or $Mg\ m^{-3}$) of soils in Lower Congo.....	202
Figure 10–6. Experimental and theoretical isotropic semi-variograms (spherical model) for bulk density of Lower Congo soils estimated by the k-nearest neighbour approach with clay, silt, sand and OC as predictors. Black circles represent the experimental semi-variance and the grey line represents the theoretical semi-variance).....	202
Figure 10–7. Spatial variability of profile water holding capacity (cm per m depth) of soils in Lower Congo.....	204
Figure 10–8. Experimental and theoretical isotropic semi-variograms (exponential model) for profile soil water holding capacity of Lower Congo soils estimated by the k-nearest neighbour approach with clay, silt, sand and pH as predictors. Black circles represent the experimental semi-variance and the grey line represents the theoretical semi-variance.	204

List of tables

Table 2-1. Classification of soils from Lower Congo datasets used to test published pedotransfer functions for soil water retention.....	17
Table 2-2. Classification of soils from Lower Congo datasets used to develop pedotransfer functions for soil water retention and bulk density.....	18
Table 2-3. Number of soil profiles and samples from the IGBP-DIS dataset per WRB Reference Soil Group used to test and/or develop pedotransfer functions for soil water retention and bulk density	19
Table 3-1. General information on point-based, parameter-based and semi-physical water retention pedotransfer functions (PTFs) developed for soils of the humid tropics.†	40
Table 3-2. Input and output variables in pedotransfer functions (PTFs) derived using data from soils in the humid tropics.†	43
Table 4-1. Representation of the 12 USDA soil textural classes in the Lower Congo dataset by decreasing order of importance	52
Table 4-2. Particle size distribution models tested in this study.....	54
Table 4-3. Initial estimates for the fitting parameters of the PSD models for 12 USDA soil textural classes	56
Table 5-1. Descriptive statistics of some physical and chemical soil properties of the Lower Congo soil dataset.†	78
Table 5-2. List of selected PTFs derived from ‘tropical’ and ‘temperate’ soils datasets.	80
Table 5-3. Soil property ranges of the calibration datasets used to develop the evaluated ‘tropical’ and ‘temperate’ pedotransfer functions (PTFs).....	84
Table 5-4. Pearson’s correlation matrix between the various potential predictors and measured soil gravimetric and volumetric water retention points ($N=60$).	85
Table 5-5. Validation indices of the gravimetric PTFs as computed on the complete evaluation dataset.†	88
Table 5-6. Validation indices of the volumetric PTFs as computed on the complete evaluation dataset.†	89
Table 6-1. Number of nearest neighbours (k) corresponding to the lowest RMSD for different values of p and different dataset sizes N_r	110
Table 6-2. Comparison of the k number generated by the power function of Nemes et al. (2006a) and the power function derived for this study.†.....	111
Table 6-3. Summary of results in terms of MD, RMSD and R^2 , for the k -NN method with optimized settings at eight different matric potentials and using 14 combinations of input attributes.†	116
Table 6-4. Prediction performance in terms of RMSD of the k -NN method using four combinations of input attributes, of the PTFs of Hodnett and Tomasella (2002) and Minasny and Hartemink (2011).†	121
Table 7-1. ‘Tropical’ and ‘temperate’ pedotransfer functions used to predict bulk density of Lower Congo soils.†	130
Table 7-2. Different levels of input variables that are used as potential predictors in the stepwise regression procedure to develop bulk density pedotransfer functions for soils in Lower Congo.†	131
Table 7-3. Different levels of input variables considered in the k -nearest neighbour algorithm to predict bulk density of soils in Lower Congo.†.....	131

Table 7-4. Pearson's correlation matrix between the various potential predictors and measured soil bulk density ($N=196$).	135
Table 7-5. Regression relationships between physico-chemical properties and bulk density for 196 soil samples from Lower Congo.†	135
Table 7-6. Prediction performance in terms of MD, RMSD and R^2 , for the k-nearest neighbour algorithm with optimized settings and using 15 levels of input attributes.†	140
Table 8-1. Different levels of availability of soil properties used as potential predictors in the stepwise regression procedure to develop water retention pedotransfer functions for soils in Lower Congo.†.	149
Table 8-2. Different levels of input variables considered in the k-nearest neighbour algorithm to predict water retention of soils in Lower Congo.†	150
Table 8-3. Pearson's correlation matrix between the various potential predictors and water content at eight different matric potentials ($N=196$).	151
Table 8-4. Coefficients of pedotransfer functions to predict water retention at eight matric potentials developed based on information on sand, silt and clay content of 196 soils from Lower Congo.†	152
Table 8-5. Coefficients of pedotransfer functions to predict water retention at eight matric potentials developed based on information on sand, silt, clay, pH and OC content of 196 soils from Lower Congo.†	152
Table 8-6. Coefficients of pedotransfer functions to predict water retention at eight matric potentials developed based on information on sand, silt, clay, pH, OC and CEC of 196 soils from Lower Congo.†	153
Table 8-7. Coefficients of pedotransfer functions to predict water retention at eight matric potentials developed based on information on sand, silt, clay, pH, OC, CEC, DCB-Fe and DCB-Al content of 196 soils from Lower Congo.†	153
Table 8-8. Coefficients of pedotransfer functions to predict water retention at eight matric potentials developed based on information on sand, silt, clay and BD of 196 soils from Lower Congo.†	154
Table 8-9. Coefficients of pedotransfer functions to predict water retention at eight matric potentials developed based on information on sand, silt, clay, BD, pH, OC and CEC of 196 soils from Lower Congo.†	154
Table 8-10. Coefficients of pedotransfer functions to predict water retention at eight matric potentials developed based on information on sand, silt, clay, BD, pH, OC, CEC, DCB-Fe and DCB-Al content of 196 soils from Lower Congo.†	155
Table 8-11. MAD, RMSD and R^2 values for hierarchical pedotransfer functions to predict water retention of soils in Lower Congo.	156
Table 8-12. MD, RMSD and R^2 values for the k-nearest neighbour algorithm with optimized settings and using 20 levels of input attributes.	159
Table 8-13. Prediction performance at various matric potentials in terms of accuracy of the MLR and the k-NN approaches developed for Lower Congo soils and in terms of reliability of the published PTFs of Hodnett and Tomasella (2002) and Minasny and Hartemink (2011) when applied on Lower Congo soils.	161
Table 9-1. Representation of the three aggregated FAO soil textural classes in the Lower Congo and the IGBP-Trop-2 datasets.	169
Table 9-2. Fitting parameter values for the cumulative PSD and the SWRC per FAO textural class for 196 soil samples of the Lower Congo dataset.†	176

Table 9-3. Fitting parameters values for the cumulative PSD and the SWRC per FAO textural class for 209 soil samples of the IGBP-Trop-2 dataset.†.....	176
Table 9-4. Performance in terms of RMSD values of the van Genuchten (1980) and three pore-solid fractal models to describe the soil water retention curve for 196 soils of the Lower Congo dataset.†.....	178
Table 9-5. Performance in terms of MD, RMSD and R^2 for the symmetric PSF model of Hwang et al. (2011) to predict water retention of Lower Congo soils (test dataset) based on the IGBP-Trop-2 dataset (calibration dataset).†.....	180
Table 9-6. Performance in terms of RMSD values for the symmetric PSF model of Hwang et al. (2011) to predict water retention of Lower Congo soils based on a k-fold cross-validation.†.....	183
Table 10-1. Site-specific and published multiple linear regression pedotransfer functions used in this study.....	191
Table 10-2. Descriptive statistics of $Clay_{avg}$, $Sand_{avg}$, OC_{avg} , BD_{avg} and PSWHC for Lower Congo soil profiles.†.....	196
Table 10-3. Prediction performance of PSWHC values by inverse distance weighting and ordinary kriging for different PTFs' approaches, models and sets of predictors through cross-validation. †....	199
Table 10-4. Prediction performance of each derived map in terms of MD, RMSD and R^2 for Lower Congo PTF-estimated PSWHC values.....	203

List of symbols

Greek symbol

Symbol	Name	Applied unit
α	scaling parameter of van Genuchten (1980) equation	kPa ⁻¹
ε	error	
γ	packing-related parameter of a soil	mm cm ⁻¹
$\gamma(l)$	semi-variance	
Ψ	soil matric potential	kPa
θ	volumetric water content	m ⁻³ m ⁻³
θ_p	volumetric water content at a predefined matric potential	m ⁻³ m ⁻³
θ_r	residual volumetric water content	m ⁻³ m ⁻³
θ_s	saturated volumetric water content	m ⁻³ m ⁻³
θ_{xkPa}	volumetric water content at x kPa matric potential	m ⁻³ m ⁻³
v	parameter related to respective shapes of particles and corresponding pores	(-)

Roman alphabet

Symbol	Name	Applied unit
a	intercept	
a	range	m
C_0	nugget	
C_0+C	sill	
C_t	soil carbon stock	kg m ⁻² 10 ⁻¹
d_{-i}	thickness of layer i	cm
d_l	size of the largest particle element	mm
d_i	size of the soil particle i	mm
d_i	distance between estimation point and observation point at location x_i	m
d_f	parameter related to the amount of fine particles	mm
d_m	diameter of the minimum allowable size particle	mm
D_p	fractal dimension of the soil water retention curve	(-)
$D_{s,B}$	fractal dimension of the cumulative particle size distribution	(-)
$D_{s,H}$	fractal dimension of the cumulative mass distribution	(-)
E_r	relative error	%
h_l	maximum matric head	cm
h_b	bubbling pressure	cm
i	individual soil sample or individual soil particle	
k	number of nearest neighbours	
K	hydraulic conductivity	mm h ⁻¹ or cm day ⁻¹
$K(h)$	unsaturated hydraulic conductivity	mm h ⁻¹ or cm day ⁻¹
K_{sat}	saturated hydraulic conductivity	mm h ⁻¹ or cm day ⁻¹
l	distance	m
m	exponent parameter in the van Genuchten (1980) equation	
n	curve shape parameter in the van Genuchten (1980) equation	
N_b	number of closest neighbours	

N_{lyr}	number of layers within 100 cm depth	
$N_p(l)$	number of data pairs separated by the distance l	
Me	moisture equivalent	$\text{kg}^{-1} \text{ kg}^{-1}$ or %
N	number of soil samples	
N_t	number of soil samples in the test dataset	
N_r	number of soil samples in the reference dataset	
pH	soil acidity in water	(-)
r	Pearson correlation coefficient	(-)
r^2	coefficient of correlation	(-) or %
R^2	coefficient of determination	(-) or %
$R^2\text{-adj}$	adjusted coefficient of determination	(-) or %
u	packing characteristic of a soil	mm cm^{-1}
w	gravimetric water content	$\text{kg}^{-1} \text{ kg}^{-1}$
W	gravimetric or volumetric water content	$\text{kg}^{-1} \text{ kg}^{-1}$ or $\text{m}^{-3} \text{ m}^{-3}$
w_i	weight assigned to soil i	
w_i	weight assigned to each $Z(x_i)$ value	
$Z^*(x_0)$	estimated value of variable Z based on the data from surrounding locations $Z(x_i)$	
$Z(x_i)$	value of variable Z at location x_i	

List of abbreviations

A-4p	four-parameter model of Andersson's
AfSIS	globally integrated-Africa soil information service
AIC	Akaike information criterion
ANN	artificial neural networks
AWC	available water capacity
BCC	bundle of cylindrical capillaries
BD	bulk density
BVTO	bias-variance trade-off
CART	classification and regression tree
CEC	cation exchange capacity
CV	coefficient of variation
DCB-Al	dithionite-citrate-bicarbonate extractable aluminium
DCB-Fe	dithionite-citrate-bicarbonate extractable iron
D.R. Congo	The Democratic Republic of Congo
ENR	extended non-linear regression
EU-HYDI	European hydropedological data inventory
F _{3p}	three-parameter model of Fredlund et al. (2000)
F _{4p}	four-parameter model of Fredlund et al. (2000)
FC	field capacity
G _{2p}	two-parameter model of Gompertz
G _{4p}	four-parameter model of Gompertz
GA	genetic algorithms
GAM	general additive model
GLM	generalized linear model
GMDH	group method of data handling
GP	genetic programming
HAC	high-activity clay
HYPRES	hydraulic properties of European soils database
IGBP-DIS	international geosphere biosphere programme data and information system
IGBP/T	subset of 'tropical' soils from the IGBP-DIS database in Hodnett and Tomasella (2002)
IGBP-Trop	subset of 'tropical' soils from the IGBP-DIS database
IGBP-Trop-2	subset of highly weathered soils from the IGBP-Trop database
INEAC	institut national pour l'étude agronomique au Congo-belge
ISRIC	international soils reference and information centre
IDW	inverse distance weighting
k-NN	k-nearest neighbour
LAC	low-activity clay
LN _{2p}	two-parameter lognormal model
LOO	leave-one-out
MAD	mean absolute difference
MARS	multiple adaptive regression splines
MD	mean difference
MIX	mixed-activity clay
MLR	multiple linear regression
OC	organic carbon

OK	ordinary kriging
OM	organic matter
POD	pore size distribution
PSD	particle size distribution
PSF	pore-solid fractal
PSWHC	profile soil water holding capacity
PTF(s)	pedotransfer function(s)
PWP	permanent wilting point
R. Congo	The Republic of the Congo
RMSD	root mean square difference
SDD	standard deviation of the differences
Sk_3p	three-parameter model of Skaggs et al. (2001)
SS	specific surface area
SSC	sand, silt, clay (soil texture)
SSE	sum of squared errors
SVM	support vector machine
SWHC	soil water holding capacity
SWRC	soil water retention curve
UNSODA	unsaturated soil hydraulic database
US	The United States
USDA	United States department of agriculture
US NRCS	US national resources conservation service
VG_2p	two-parameter model of Haverkamp and Parlange (1986)
VG_3p	three-parameter model of Haverkamp and Parlange (1986)
W	gravimetric or volumetric water content
W_3p	three-parameter model of Weibull

1 General introduction

1.1 Background and nature of the problem

Ensuring food security is one of the main challenges developing countries in the tropics are facing. This is particularly true for Sub-Saharan Africa which is the only region where per capita food production has remained stagnant over the past 40 years (Pinstrup-Andersen et al., 1999). Depletion of soil fertility along with the concomitant problems of weeds, pests, and diseases, shortage of water for agriculture due to climate change and land degradation, increasing costs of irrigation systems associated with the rising price of fuel, soil and groundwater pollution due to continuous application of fertilizers and pesticides are amongst the major causes of hunger and poverty in various tropical regions. This is particularly true for Africa where agriculture represents the economic sector that engages 70% of the population. Food insecurity in Africa is directly related to insufficient total food production (Sanchez, 2002).

The productivity of most soils in the tropics is mainly limited by physical and chemical constraints and to some extent by climatic constraints. According to Von Uexkull and Mutert (1995), the high prevailing temperatures and the resulting rapid chemical, physical and biological processes mean that the penalties paid for imbalances in the solid, liquid and gaseous phases of soils in tropical regions are much larger than in temperate regions. These problems can be solved and consistently high agricultural yields obtained only by the development and introduction of comprehensive management packages (Yerima and Van Ranst, 2005). Therefore there is an urgent need to increase food production and reduce poverty through better land management policies.

The unsaturated (vadose) zone of the Earth attracts a substantial and increasing attention because of the negative impacts of human activities on the quality of the subsurface environment. Sustainable development can only be achieved through a rational management of natural resources. This is an important issue for developed countries as well as for developing countries, such as the Democratic Republic of Congo (D.R. Congo).

Soil and water constitute two important natural resources. They are fundamental supports of natural and artificial ecosystems and human life. Soil and water are intimately related in nature. Therefore, a deep understanding of their properties, behaviours and interactions is essential for their rational management. This requires a collaborative effort of various soil and water related disciplines and expertises.

In a paper discussing the changing paradigm of agricultural research in relation to precision agriculture and pedology, Bouma et al. (1999) indicated that new interdisciplinary approaches are needed to improve the understanding of complex interactions between multiple factors affecting crop production and farm decision making. There is growing recognition that synergy can be generated by bridging traditional pedology with soil physics and hydrology to enhance integrative studies of soil-water relationships across spatial and temporal scales (Lin, 2003).

By integrating different soil-related fields as shown in Figure 1–1, hydropedology can address: (1) knowledge gaps between pedology and soil physics/hydrology; (2) multiscale bridging from microscopic to mesoscopic and to macroscopic levels; and (3) data translations from soil survey databases into soil hydraulic information (Lin, 2003). The emerging interdisciplinary research field of hydropedology attracts substantial attention because of its promise to bridging pedology and hydrology (Pachepsky et al., 2006). Hydropedology addresses soil hydrologic functioning at various scales by applying data fusion, PTFs, and concurrent use of models (Pachepsky et al., 2008).

Soil hydraulic properties are physical properties that describe the soil-water relationship. Soil hydraulic properties, such as water retention and hydraulic conductivity (K), are of utmost importance in soil related fields such as hydrology, agronomy, meteorology, or contaminant hydrology and geochemistry (Rawls et al., 1991), just to mention a few. Other soil physical properties such as bulk density are needed for studies

related to estimation of carbon stocks and nutrient status of soils (Bernoux et al., 1998). The soil water retention curve (SWRC) and the soil bulk density (BD) will be designated in this dissertation as the hydrophysical properties of the soil.

Direct measurements of soil hydrophysical properties are still laborious, expensive and time-consuming (Harrison and Bock, 1981; Klute, 1986; Dirksen, 1991) especially for environmental studies at a regional scale (Vereecken, 1995; Basile and D'Urso, 1997; Pachepsky et al., 2006; Guber et al., 2006; Minasny and Hartemink, 2011; Wösten et al., 2013). Particularly in developing countries, there are additional problems associated with this task, ranging from personnel training to acquisition of the necessary equipment (Medina et al., 2002). Therefore, an attractive alternative to the direct and often cumbersome measurements of soil hydrophysical properties is their estimation by so-called pedotransfer functions (PTFs).

Bouma (1989) described the term pedotransfer function (PTF) as 'translating data we have into what we need'. Hydraulic PTFs are predictive functions which relate more easily measurable soil data such as soil texture (sand, silt and clay), bulk density, organic matter (or organic carbon) content and/or other data routinely measured or registered in soil surveys with hydraulic parameters such as SWRC (Bouma and van Lanen, 1987; Bouma, 1989; van den Berg et al., 1997). Estimating soil hydraulic characteristics from readily available physical parameters has been a long-term goal of soil physicists and engineers (Saxton and Rawls, 2006), and is still one of the hot topics in soil science. McBratney et al. (2011) wrote that PTFs have become an active soil science research topic with an increasing number of papers and citations since the term was coined more than 20 years ago by Johan Bouma (Bouma, 1989). Recently, at the 4th International Congress of the European Confederation of Soil Science Societies (EUROSOIL 2012), an entire session dedicated to pedotransfer functions was organized with the title: 'Pedotransfer functions in soil hydrology: still a myth, or a pleasant reality?: applications, validations, and case studies'. Various researchers discussed and outlined perspectives for research in pedotransfer functions. In its introductory presentation entitled: 'Applying pedotransfer functions: the sorcerer's apprentice at work?', Johan Bouma stated that PTFs are a nice illustration of hydropedology at work.

During the last decades, many PTFs have been developed to predict SWRC, K and BD. However, as observed by Tomasella et al. (2000), Hodnett and Tomasella (2002) and Tomasella and Hodnett (2004), most PTFs available in the literature have been derived and tested using extensive soil databases of temperate regions. Schaap (2004) noticed that current international databases have a serious bias towards soils from temperate regions and very little data is available for 'tropical' soils. In a similar vein, Schaap (2005) wrote that 'with the exception of a few studies, hydraulic data and corresponding indirect methods about tropical soils are a virtual *terra incognita*'. As the use and development of PTFs progressed, several problems related to the applicability of PTFs developed from temperate-climate soils datasets when applied to tropical-climate soils datasets arose and were articulated by several researchers (e.g., Van Ranst, 1995; van den Berg et al., 1997; Medina et al., 2002; Hodnett and Tomasella, 2002; Tomasella and Hodnett, 2004; Minasny and Hartemink, 2011).

Van Ranst (1995) indicated that with increasing interest in development of agricultural land in the humid tropics in recent years, has come the realization that certain soils which are often red or yellow have some unique physical and chemical properties that distinguish them from soils commonly found in temperate regions. Soil mineralogy is a key factor controlling the physical and chemical properties important for managing these soils. Marked differences in physical, chemical and mineralogical properties between soils from temperate regions and soils from the humid tropics might explain why the PTFs derived for soils of Europe and North America seem not to be suitable for highly weathered soils of Sub-Saharan Africa and South America. Various studies (e.g. Sharma and Uehara, 1968; Correa, 1984; Demattê, 1988; Tomasella and Hodnett, 1998; van den Berg et al., 1997; Hodnett and Tomasella, 2002; Minasny and Hartemink, 2011) showed clearly that soils of the humid tropics present different water retention properties compared to their

temperate region counterparts with low BD ($0.7\text{--}1.2 \text{ Mg m}^{-3}$), high permeability (saturated K usually ranging from 10 to 1000 mm h^{-1}), low available water capacity, AWC (70 mm m^{-1}) and approximately 80% of the plant available water between -10 and -100 kPa .

Until now, only a limited number of studies has been carried out with international databases of soils from the humid tropics (Tomasella and Hodnett, 2004; Minasny and Hartemink, 2011) compared to soils from temperate regions. Recently, Minasny and Hartemink (2011) indicated that fewer efforts are devoted in prediction of properties of soils of the tropics where the need for accurate and up-to-date soil properties information is even more urgent because soil data in these regions are often sparse and outdated. Consequently, there exist a certain number of knowledge gaps related to pedotransfer modelling of hydrophysical properties of soils of the humid tropics and its potential applications for land management.

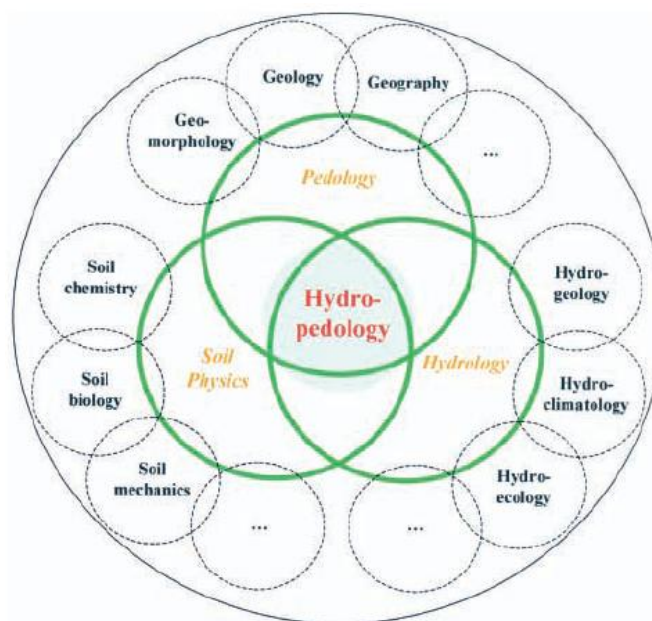


Figure 1–1. A conceptual diagram showing the relationships of hydropedology to other related disciplines (from Lin, 2003)

1.2 Knowledge gaps

Most African countries located in the humid tropics lack information on soil hydrophysical properties in their databases (Leenaars, 2013). Lower Congo located in the South-western part of the D.R. Congo is a striking illustration of this paucity. For Lower Congo, a large soil database of various physico-chemical properties of more than 1,400 soil horizons representing more than 300 soil profiles is available (Baert et al., 1991a; Baert, 1995). This region has been an active research area in the field of soil science and soil related disciplines since the 1940s. This resulted in a comprehensive soil database, completed with climatological and agricultural data. However, soil hydrophysical data like SWRC, K and BD are lacking. This is principally due to the fact that previous soil survey campaigns and physico-chemical analyses in the D.R. Congo were undertaken to characterize, classify and map ‘tropical’ soils, to determine their chemical fertility and to conduct land evaluation studies (Baert et al., 2012).

There is an urgent need to provide reliable information on these hydrophysical properties through their estimation by PTFs. To date, no hydrophysical PTFs have been developed for soils from the D.R. Congo in general and for soils in Lower Congo in particular. In the literature, there exists a limited number of PTFs developed for soils of tropical regions, though it has not been investigated so far whether they are applicable

to Lower Congo. It is neither unknown which predictor variables would be most appropriate to explain variation in hydrophysical characteristics such as SWRC and BD in Lower Congo.

During the early days of data collection in soil science and in soil hydrology, scientific investigations were typically limited to single studies, or at best studies that relied on data collections of single research institutions. It has been recognized already decades ago that there is great value in assembling larger, national and international data collections. For the humid tropics, a number of national and regional databases exists, as well as there is an increase of international data collections that contain soil hydraulic data, though quite limited in number, size and complexity. Nowadays, efforts are made for exchange and extension of databases in order to develop PTFs at national, regional and continental scales. The most recent attempt in joining national datasets was initiated in 2012 by the Joint Research Centre of the European Union at Ispra, with the objective of collecting hydraulic and basic soil properties from EU member states (EU-HYDI; European Hydropedological Data Inventory). The compiled database will then be used to develop European PTFs. Previous examples include UNSODA, HYPRES, IGBP-DIS.

To develop hydrophysical PTFs at different scales, texture (i.e. clay, silt, sand fractions or more detailed fractions of soil particles) is generally considered as the minimum basic soil information that is required. However, in many countries in Africa (like in other continents), different particle size fractions and classes have been used in their national classification systems for soil texture. To achieve compatibility between different national databases, continuous particle size distribution (PSD) models that enable the conversion from one textural system to another one are therefore needed (Nemes et al., 1999). However, very few and limited studies have been conducted to estimate the performance of existing PSD models when applied to soils of the humid tropics.

On the other hand, those PTFs that were developed for soils of humid Africa are mostly based on multiple linear regression (MLR) equations. Despite the limited efforts in data collection and harmonization for soils of the humid tropics (where most of the developing countries are located) compared to temperate areas, large tropical soil databases such as the African Soil Profiles Database 1.1 are expected to steadily grow (Leenaars, 2013). With the emergence of such large databases and with increasingly complex applications in view, classical statistical methods such as MLR may have limitations because important trends may not be detected. Moreover, as new data will become available, regression PTFs should be redeveloped (Nemes et al., 2006a) if those data are to be included in the PTF. Better and more flexible tools to analyse soil databases and to provide necessary soil data to emerging applications are therefore warranted. Data mining techniques, e.g., k-nearest neighbour (k-NN) can help recognize data patterns without assuming any underlying model equations. In recent studies, the k-NN has been found to be a competitive alternative to classical equation-based PTFs to predict water retention at -33 -kPa and -1500 -kPa matric potentials (Nemes et al., 2006a,b). Studies on the potential of this pattern-recognition technique in prediction of water retention have been limited to soils from the temperate areas. To our knowledge, there are no studies which applied the k-NN technique to estimate water retention of highly weathered soils in the humid tropics. Additionally, there are no studies which compared the performance of equation-based PTFs and pattern recognition-based PTFs in their estimation of water retention of highly weathered soils.

Allen and Hajek (1989) stated that Fe oxides are the most abundant metallic oxides in soils and they occur in most soils. Pedomorphically, Fe oxides are perhaps one of the most important components of 'tropical' soils according to Schwertmann and Taylor (1989). In their first studies dedicated to pedotransfer modelling of 'tropical' soils, dithionite-citrate-bicarbonate extractable iron and aluminium (DCB-Fe and DCB-Al) have been used with success by van den Berg (1996) and van den Berg et al. (1997) in predicting water retention of strongly weathered soils. However, no additional investigations have been conducted to confirm the potential of these chemical soil properties in improving the prediction of hydrophysical properties of soils of the humid tropics. One of the reasons might be that information on Fe and Al contents is poorly available in

soil survey databases. Since DCB-Fe and DCB-Al data are present in the Lower Congo database, it is therefore worthy to investigate their relative importance as potential predictors of water retention of highly weathered soils in the humid tropics.

Although statistical data mining tools based on a variety of predictor variables are most widely used for developing soil hydraulic PTFs, a number of studies have been devoted to semi-physical models (Arya and Paris, 1981; Haverkamp and Parlange, 1986; Tyler and Wheatcraft, 1990; Bird et al., 2000). They recognize the shape similarity between the cumulative PSD and soil water retention characteristics. Despite their limited applicability in pedotransfer modelling, they do offer valuable conceptual insights into the physical relations between PSD and pore size distribution (POD). In the characterization of soil water retention by semi-physical models, the fractal approach based on pore-solid fractal (PSF) models has gained attention in the modelling community due notably to the growing interest for fractal theory in soil science. At present, there are only a limited number of studies related to the application of PSF models in the description and prediction of the water retention parameters. Moreover, the few examples of applications of PSF models have been limited to temperate-climate soils (Bird et al., 2000; Hwang et al., 2011) and recently to dry-climate soils (Bayat et al., 2013). Lack of detailed PSD data has been pointed as one of the constraints to applying the fractal approach in pedotransfer modelling of water retention (Cornelis et al., 2001; Khlosi, 2003). The potential of PSF models to describe and predict water retention characteristics of highly weathered soils of the humid tropics has not yet been investigated. Yet, availability of detailed PSD in some large tropical soil databases makes the implementation of such investigation possible.

Characterization of BD and soil water holding capacity (SWHC) is important for assessing the soil's physical status and quality. Assouline (2006) wrote that an increase in BD influences many aspects of the soil–water–plant–atmosphere system. The influence of many soil management practices on soil properties that affect water movement and water retention in soils can be reflected in the soil BD (Kaur et al., 2002). Furthermore, soil OC estimates require data on BD which serve as a conversion factor. On the other hand, SWHC is one of the most important soil factors for plant growth, influencing carbon allocation, nutrient cycling, and the rate of photosynthesis (Milly and Shmakin, 2002; Piedallu et al., 2011; Hong et al., 2013). As indicated by Piedallu et al. (2011) and Hong et al. (2013), little research has focused on mapping the SWHC despite a clear interest of soil hydraulic parameters for climatological, ecological and hydrological studies. This is particularly true for tropical zones of Sub-Saharan Africa such as Lower Congo where data scarcity on soil hydrophysical characteristics namely BD, SWRC and K is even bigger than for temperate regions (Batjes, 2008; Wösten et al., 2013). To date, a BD and a SWHC maps of Lower Congo based on soil information collected from various soil survey campaigns and using site-specific and published PTFs are still missing.

1.3 Objectives of the study

The overall objective of this work is to test and develop indirect methods for predicting hydrophysical properties of soils in the Lower Congo region.

The specific objectives of this study are:

- to establish the first extended dataset on hydrophysical properties of soils in Lower Congo;
- to evaluate the performance of models for describing cumulative particle size distribution (PSD) of soils in the humid tropics in order to achieve compatibility between soil databases from various tropical regions for multiscale pedotransfer function (PTF) modelling;
- to evaluate the performance of published PTFs to predict water content of soils in Lower Congo;
- to apply, enhance and test the pattern recognition technique k-nearest neighbour (k-NN) to obtain estimations of water content at various matric potentials of soils from the humid tropics;
- to develop equation-based PTFs and compare them with the pattern recognition-based (k-NN) PTFs for predicting hydrophysical properties of soils in Lower Congo;

-
- to assess the ability of physico-empirical models based on the pore-solid fractal (PSF) theory to describe and predict water retention of highly weathered soils;
 - to construct a BD and a soil water holding capacity (SWHC) maps as a support to decision makers for a better land resources management of the Lower Congo region.

1.4 Outline of the dissertation

This dissertation is structured around the specific objectives outlined above and is divided in eleven chapters preceded by a short introduction. Chapter 2 gives a general description of the physical and the economic environment of the study area. Chapter 3 presents an overview on water retention PTFs developed for soils of the humid tropics. Chapter 4 deals with the evaluation of fitting performance of PSD models for soils of the humid tropics based on a large dataset of soils from Lower Congo. Chapter 5 presents an evaluation of already published PTFs to predict water retention properties of soils in the humid tropics. Chapter 6 deals with the application, enhancement and test of the non-parametric k-NN approach to predict water retention of soils in the humid tropics based on pattern-recognition. Chapter 7 treats the development of equation-based and pattern recognition-based PTFs to predict BD for soils of Lower Congo and the identification of the best predictors. Chapter 8 reports the development of equation-based and pattern recognition-based PTFs as well as the identification of the best predictors of water retention at various matric potentials for soils of Lower Congo. Chapter 9 focuses on the application of physico-empirical models based on the pore-solid fractal theory to describe and predict water retention of highly weathered soils. Chapter 10 illustrates the practical application of water retention PTFs in the construction of a BD map and a SWHC map of Lower Congo as a support for a better land management of this region. Chapter 11 summarizes the key findings of our research and gives some recommendations for future research.

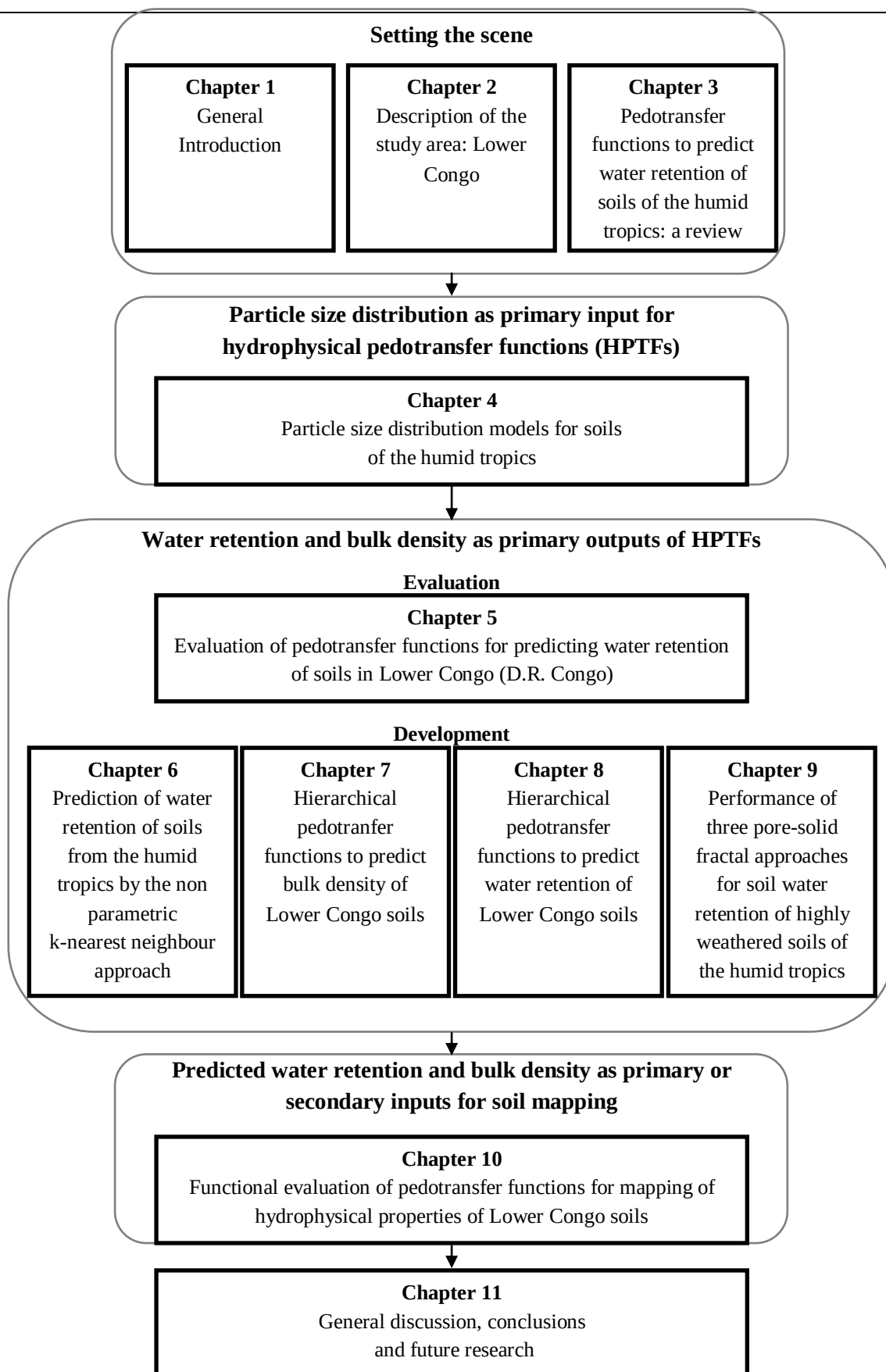


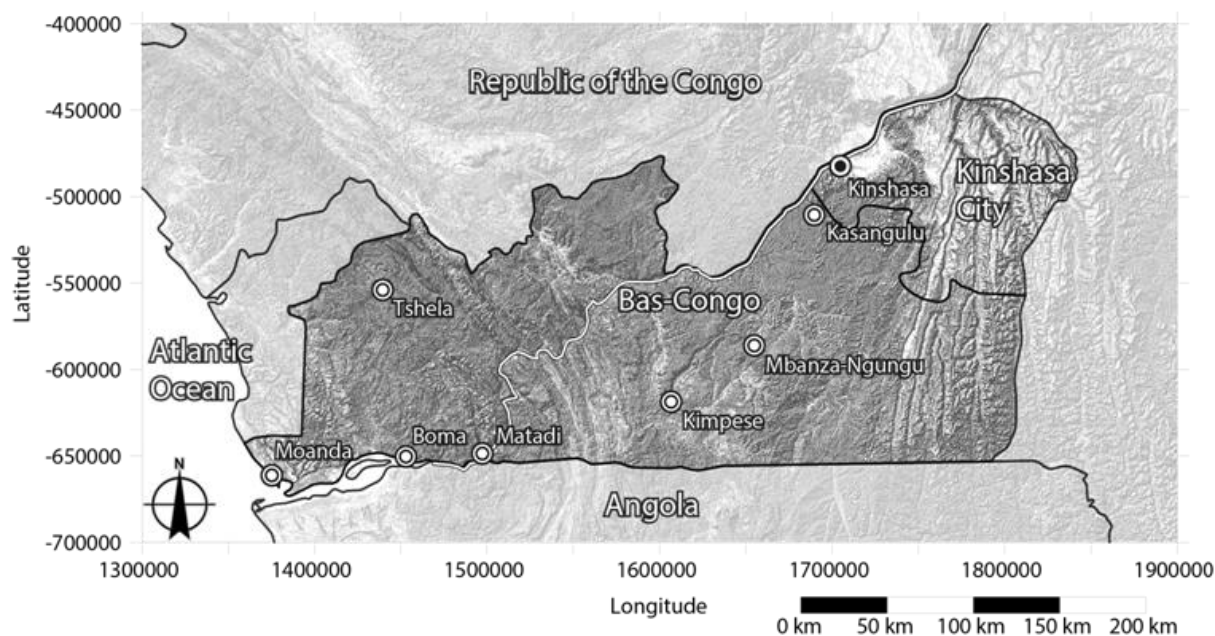
Figure 1–2. Schematic overview of the dissertation.

2 Description of the study area: Lower Congo

2.1 Biophysical setting

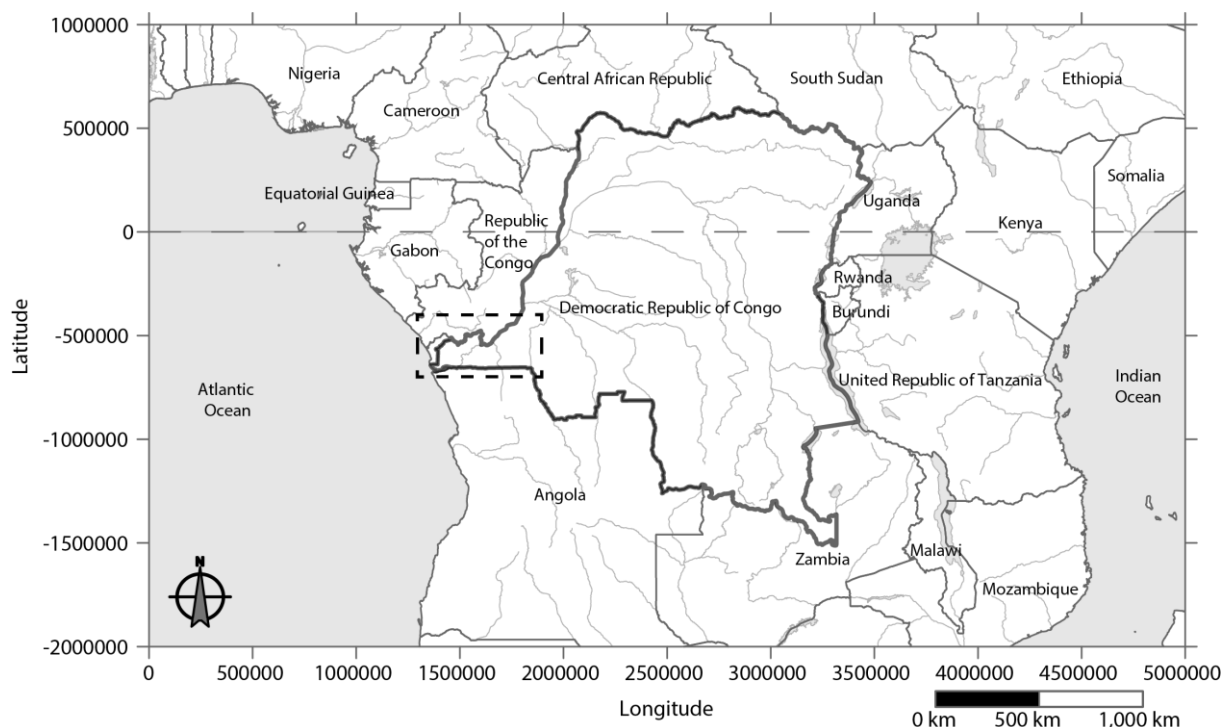
2.1.1 Geographical location

The term ‘Lower Congo’ is used to indicate the western part of the D.R. Congo that is located between the Congo River and the Atlantic Ocean (Figure 2–1). Lower Congo lies between the latitudes 4° and 5° S and the longitudes 13° and 16° E, covering a total area of almost 64,000 km². It is bordered by Angola to the South, by Cabinda and the Republic of Congo (R. Congo) to the North, by the Atlantic Ocean to the West and by the province of Bandundu to the East (Figure 2–2). Lower Congo comprises two important administrative entities: the province of Bas-Congo and the capital city of Kinshasa.



Source: Jan De Pue, Soil Physics Research Unit, Ghent University. Software used: Surfer v11.1.719 (Golden Software, Inc., Golden, CO).

Figure 2–1. Lower Congo is the South-western part of the D.R. Congo located between the Congo River and the Atlantic Ocean and comprises the province of Bas-Congo and the capital city of Kinshasa.



Source: Jan De Pue, Soil Physics Research Unit, Ghent University. Software used: Surfer v11.1.719 (Golden Software, Inc., Golden, CO).

Figure 2–2. Lower Congo is bordered by Angola to the South, by Cabinda and the Republic of Congo to the North, by the Atlantic Ocean to the West and by the province of Bandundu to the East.

2.1.2 Climate

The atmospheric climate of Lower Congo belongs to the Aw_4 and Aw_5 type according to the Köppen climatic classification system (Bultot, 1950). It is a humid, tropical climate with a marked dry season of four to five months, from June to September, during which a clear decrease in temperature can be observed. Annual precipitation ranges from 1,200 mm (Bas-Fleuve district) to 1,400 mm (Cataractes district). The annual average temperature is around 25° C. During the dry season, there is almost no rainfall and the temperature is around 22° C (PNUD/UNOPS, 1998a,b). The soil climate influences the pedogenetic processes, the vegetation growth and land use. In most classification systems, the pedoclimate is used at very high level. The USDA Soil Taxonomy distinguishes the ‘Soil Temperature Regime’ and the ‘Soil Moisture Regime’ (Soil Survey Staff, 1975). The temperature regime is classified as Isohyperthermic, i.e. the mean annual soil temperature is 22° C or higher. Most soils in Lower Congo (apart from the poorly drained soils with an Aquic soil moisture regime) have a Ustic moisture regime (Udic Tropustic) (Baert, 1995).

2.1.3 Geology

In Lower Congo, numerous geological surveys have been carried out and different geological maps (published and unpublished) have been prepared. Lower Congo consists of a Precambrian basement, covered over vast areas by Mesozoic sediments. The physiography is related to the repartition of the geological formations. Three distinct geological and geomorphological regions can be distinguished, which from West to East are: the coastal region, the axial region and the oriental region (SOGREAH–SOCINCO, 1964; Stoops, 1966). The interested reader can refer to Baert (1995) for further information.

2.1.4 Topography

The landscape of Lower Congo is very diversified as a result of the various lithologies of the geologic materials and the succession of erosion cycles (Steenstra, 1988). Several erosion cycles have occurred since the Cretaceous, which resulted in a series of, often laterized, peneplains (Cahen and Lepersonne, 1948; Nicolai, 1961). There is a strong relationship between the present landforms and geological structure, and the differences in the weathering resistance of the various rock types in Lower Congo. As a general rule, the resistant rocks correspond to the higher parts in the landscape, while the less resistant rocks coincide with the lower parts.

2.1.5 Vegetation

The actual vegetation in Lower Congo is dominated by an open shrub savannah, mainly anthropic and preserved due to inappropriate cultural practices: deforestation, shifting cultivation with too short fallow periods, lack of soil conservation practices and annual bush fires (Devred, 1956; Compere, 1970). However, the climax vegetation in most of Lower Congo seems to be a dense forest (Duvignaud, 1949; Maudoux, 1954; Mahieu, 1961). The very sandy soils of the coastal region in Bas-Congo and the Bateke Plateaus in Kinshasa are covered by a steppe vegetation (Duvignaud, 1949; Baert et al., 1991b).

2.2 Soils

2.2.1 Previous pedological studies in Lower Congo

The first pedological study in Lower Congo was carried out in the 1930s (Baeyens, 1938). The principal aim of this study was to determine the natural fertility of the soils. The interest in soils of Lower Congo strongly increased from 1940 onwards with the installation of several agricultural research stations by the former 'Institut National pour l'Etude Agronomique au Congo-belge, INEAC' (in Luki, Kondo, Gimbi and M'Vuazi) and by 'Groupe d'Economie Rurale, GER' (in Mbanza Ngungu, Luala and Kimpangu). Soil research and soil surveys were mainly concentrated around those stations.

The soil map of INEAC station at M'Vuazi (Denisoff and Devred, 1954) was published at scale of 1:50,000. In 1958, the Division for Soil and Botanic Survey of INEAC started a systematic reconnaissance soil survey of Lower Congo at scale 1:200,000 in order to locate sites suitable for agricultural development. Unfortunately, due to the events related to the independence of Congo (which was from 1971 to 1997 called Zaïre), this survey, carried out under the supervision of the late Carolus Sys, was never terminated. The legend of the soil maps was based on the INEAC classification (Sys et al., 1961) for soils developed on different parent materials, under forest or savannah, with supplementary information on erosion and degradation, landforms and sometimes soil texture.

During the systematic INEAC soil survey, several supplementary soil surveys of agricultural sites were carried out by private firms (SOCINCO, 1959a,b; SOCINCO, 1960; SOGREAH, 1959) which were still carried out after independence in order to relaunch the agricultural activities (e.g., SOGREAH-SOCINCO, 1964; IFAGRARIA, 1975; Brouwers, 1987). Generally, the soil maps only indicate differences in texture, drainage and soil depth. Jost (1977) compiled the existing soil data into a very general soil map of the Lower Zaïre (former name of Lower Congo), later published in the Atlas of the Lower Zaïre (BEAU, 1987).

2.2.2 Systematic soil survey in Lower Congo

Under the impulse of Carolus Sys, the systematic soil survey of Lower Congo at reconnaissance level was relaunched in 1988. It was a three-year project (1988–1990) financed by the Belgian General Administration for Development Cooperation (GADC). The principal objectives of this project were: (1) a soil survey of the whole Lower Congo at a scale of 1:200,000; and (2) an inventory of the different soil types. Baert et al. (1991b) elaborated land suitability maps which are useful to select suitable areas for future agriculture development. The reconnaissance survey of Lower Congo region was a free survey. The observations were carried out in profile pits, in artificial road cuts and trenches, in quarries, in natural erosion gullies and ravines and by augering. More details on the history of soil survey in Lower Congo can be found in Baert (1995).

Soils in Lower Congo are highly weathered and do not behave as soils in temperate areas due to marked differences in their mineralogical composition. ‘Temperate’ soils are characterized by a predominance of permanent charge minerals in their clay fraction while ‘tropical’ soils predominantly contain variable charge minerals, e.g., kaolinite and sesquioxides. Regarding their mineralogy, soils from the humid tropics, as those in Lower Congo, present unfavourable properties for agricultural use: low effective cation exchange capacity, low nutrient holding capacity and reserve, low available water holding capacity, high K, high phosphorus adsorption and high point of zero charge (Van Ranst, 1991, 1995).

Therefore, these soils need specific management practices to improve their productivity which is generally very low for most cultivated plants. Improvement of the fertility level of soils in Lower Congo requires a good understanding of the behaviour of variable charge colloids in these soils. For that purpose, Baert (1995) selected 17 profiles, derived from different parent rocks for a more detailed study. Additionally, Marcelino (1995) studied the characteristics and genesis of soils in sandy materials. Koy (2005) was interested in determining the impact of a planted *Acacia* forest on chemical fertility of some sandy soils in the Bateke Plateaus. In the frame of the underlying dissertation, a new field campaign was initiated and undisturbed soil samples on 24 profiles were collected from various locations of the Lower Congo region.

Lower Congo is one of the regions of the D.R. Congo which benefited of the most intensive soil survey campaigns. As previously mentioned, more than 300 soil profiles were described for this region with more than 1,400 soil samples which were analysed for various physico-chemical properties at the Laboratory of Soil Science at Ghent University (Belgium). More precisely, 1,412 soil samples collected during the systematic soil survey of Lower Congo with information on detailed PSD were selected to test ten different PSD models (Chapter 4).

However, data on hydrophysical properties (SWRC, K and BD) are dramatically lacking. This can be considered as the weakest point of the Lower Congo database as a support for a better land management of this region through improvement of soil physical fertility.

2.2.3 Soil samples used for pedotransfer modelling

2.2.3.1 Soils selected from Lower Congo datasets

The 16 studied soils used in the preliminary evaluation study on published PTFs reported in Chapter 5 were described and classified by Baert (1995), Marcelino (1995) and Koy (2005) who worked on various aspects of physico-chemical and mineralogical properties of soils of Lower Congo. The interested reader can refer to the respective doctoral theses for more information. In this study, 24 soil

profiles provided 196 soil samples that were collected in the autumn of 2009 and that were used for pedotransfer modelling. From this dataset, 139 soil samples were selected and used as an independent evaluation dataset to assess the ability of a non-parametric method to predict water retention of soils of the humid tropics (Chapter 6). In Chapters 7 and 8, all the 196 soil samples were used as development dataset to formulate water retention and BD PTFs for the Lower Congo region. The 24 soil profiles were classified according to IUSS Working Group WRB (2006). From Table 2–1 and Table 2–2, it can be seen that Ferralsols, Acrisols and Arenosols are the dominant WRB Reference Soil Groups. More details on the soil samples collected from the 24 soil profiles will be given in the subsequent chapters.

Table 2-1. Classification of soils from Lower Congo datasets used to test published pedotransfer functions for soil water retention

Soils used to test published pedotransfer functions	
WRB Reference Soil Groups (FAO-UNESCO-ISRIC, 1990)	Profile name
Haplic LUVISOL	SB34
Humic ACRISOL to Rhodic FERRALSOL	BO5
Geric FERRALSOL	SB31
Luvic PHAEZEM	SO31
Xanthic FERRALSOL	SO13
Haplic FERRALSOL	MG62
Haplic ACRISOL	BO53
Haplic ACRISOL	MT18
Xanthic FERRALSOL	MT2
Dystric CAMBISOL	TSH47
Humic ACRISOL	SB41
Haplic LIXISOL	BO52
Geric FERRALSOL	KIN25
ARENO-FERRALSOL	KIN24
Carbic PODZOL	KIN7
Rubic Ferralic ARENOSOL	KIN1

Table 2-2. Classification of soils from Lower Congo datasets used to develop pedotransfer functions for soil water retention and bulk density

Soils used to develop pedotransfer functions	
WRB Reference Soil Groups (IUSS Working Group WRB, 2006)	Profile name
Nitic ACRISOL (Epieutric, Clayic) to Mollic NITISOL (clayic)	BOM1
Umbric FERRALSOL (Dystric, Clayic)	GIM1
Umbric FERRALSOL (Dystric, Clayic)	GIM2
FERRALSOL (Dystric, Clayic)	GIM3
FERRALSOL (Dystric, Clayic)	KING1
Umbric FERRALSOL (Dystric, Clayic)	ING1
Umbric FERRALSOL (Dystric, Clayic)	MAT1
Umbric FERRALSOL (Dystric, Clayic)	MAT2
Umbric FERRALSOL (Dystric, Clayic)	LUT1
Ferralic ARENOSOL (Dystric)	LUT2
Umbric NITISOL (Dystric)	KIM1
Umbric FERRALSOL (Dystric, Clayic)	KIMV1
Umbric ACRISOL (hyperdystric, Clayic)	LUK1
Umbric ACRISOL (Ferric, Dystric)	MVZ1
Umbric FERRALSOL (Dystric, Clayic)	MVZ2
Umbric NITISOL (Dystric)	MVZ3
Umbric FERRALSOL (Dystric, Clayic)	KWI1
Umbric FERRALSOL (Dystric, Clayic)	KWI2
ARENOSOL (Dystric)	NDJ1
Ferralic ARENOSOL (Dystric)	NDJ2
ARENOSOL (Dystric)	NDJ3
ARENOSOL (Dystric)	NDJ4
Ferralic ARENOSOL (Dystric)	NDJ5
ARENOSOL (Dystric)	NDJ6

2.2.3.2 Soils selected from the international IGBP-DIS dataset

In this study, soil data were also taken from the large international database of the International Geosphere Biosphere Programme Data and Information System (IGBP-DIS) from the International Soil Reference and Information Centre (ISRIC) in Wageningen, the Netherlands (Tempel et al., 1996). The soil profiles selected for this research work were essentially located in tropical regions. Tropical regions here are defined as regions situated between 25° N and 25° S and mainly under a (sub)humid climate. Soils within the tropics but under temperate climate conditions due to altitude or in dry areas were not included in the selected dataset. This ‘tropical’ subset of the IGBP-DIS dataset containing 534 soils samples is referred to here as the ‘IGBP-Trop’ dataset. For sake of parsimony, only the WRB Reference Soil Groups of the 534 soil samples constituting the IGBP-Trop subset are listed in Table 2–3. The IGBP-Trop dataset was used as training (reference) dataset in Chapter 6. A part of the IGBP-Trop dataset containing only Ferralsols, Acrisols and Arenosols was designated as the ‘IGBP-Trop-2’ dataset. This subset of 209 soil samples was used in Chapter 9 as an independent calibration dataset.

More information about the IGBP-DIS international database can be found on the ISRIC World Soil Information website (<http://www.isric.org/data/igbp-dis-soil-data-set-pedotransfer-function-development>, consulted on 24th April 2013). Further details on the selected soil samples constituting the IGBP-Trop subset will be given in the following chapters.

2.2.4 Soil samples used for mapping soil hydrophysical properties in Lower Congo

In Chapter 10, two datasets based on soil samples collected within 100 cm depth on the 24 soil profiles (155 soil samples) from the autumn 2009 fieldwork campaign and on 248 profiles (972 soil samples) from the systematic soil survey of Lower Congo have been used to derive maps of soil hydrophysical properties. More details on the use of these two soil datasets for mapping purposes will be given in Chapter 10.

Table 2-3. Number of soil profiles and samples from the IGBP-DIS dataset per WRB Reference Soil Group used to test and/or develop pedotransfer functions for soil water retention and bulk density

WRB Reference Soil Groups (IUSS Working Group WRB, 2006)	Number of soil profiles	Number of soil samples
Acrisols	25	62
Andosols	14	34
Arenosols	7	13
Cambisols	25	76
Chernozems	3	12
Ferralsols	39	109
Fluvisols	4	15
Gleysols	9	25
Kastanozems	1	2
Luvissols	13	32
Nitisols	8	25
Phaeozems	8	23
Planosols	3	5
Podzols	1	4
Regosols	3	11
Solonchaks	1	3
Solonetz	2	4
Vertisols	7	14
Undefined	23	65

2.3 Socio-economic setting

As mentioned above, Lower Congo is composed of two provinces: the province of Bas-Congo and the city-province of Kinshasa. The province of Bas-Congo is the only province with a coastline (Atlantic Ocean) and has an area of about 53,920 km². It is one of the most active provinces of the D.R. Congo with a very developed economy: industrial crops (rubber, cocoa, robusta coffee, oil palm, sugar cane), food crops (banana, cassava, maize, sweet potato, yam, cajan pea, taro, groundnuts, beans, onions, pepper), mineral products (oil), industries (sugar refinery, flour milling, oil-works, cement factory, oil refinery) among other things. Various economical activities were made possible thanks to several assets: particular geographical location (Atlantic coastline and proximity with the capital city), forest biodiversity, arable lands, a railway linking the capital city Kinshasa and the port city of Matadi, a bridge, the 'pont Matadi' formerly known as 'Pont Maréchal Mobutu', linking the port cities of Boma and Matadi and giving access to the Moanda beach, the two hydroelectric dams of Inga, among others.

For these last reasons as well as for other geographical and historical reasons, Bas-Congo is one the most touristic provinces of the D.R. Congo (Wikipedia, 2013a, Tombu and Huart, 2012).

Kinshasa is the capital city of the D.R. Congo and covers an area of 9,965 km². It is the largest and the most populated city of the D.R. Congo with a population estimated to more than nine million. Kinshasa is a city of sharp contrasts. Over 90% of the city-province's land is rural in nature, and the urban area only occupies a small section in the far western end of the city-province. The food supply of this densely populated city is entirely dependent on the food production of its agricultural (rural) zones and neighbouring provinces such as Bas-Congo and Bandundu. With a flourishing community of scientists, musicians and artists, Kinshasa can be considered as a major intellectual and cultural centre in Central Africa. It is also the country's major industrial centre, processing many of the natural products brought from the hinterland (Wikipedia, 2013b). The interested reader can refer to monographic series of PNUD/UNOPS (1998a,b) on the provinces of Bas-Congo and Kinshasa for more detailed information.

3 Pedotransfer functions to predict water retention of soils of the humid tropics: a review

This chapter is based on Botula, Y.-D., Van Ranst, E., Cornelis, W.M., 2013. Pedotransfer functions to predict water retention of soils of the humid tropics: a review (article under revision in Revista Brasileira de Ciencia do Solo).

3.1 Introduction

Soil hydraulic properties, e.g., water retention, are cumbersome, time-consuming and costly to measure and they also change over time. Therefore, soil scientists and hydrologists have searched alternative methods for a fast and accurate prediction of difficult-to-measure soil properties. During the last three decades, estimation methods, called pedotransfer functions (PTFs) have been largely used by soil scientists in temperate regions to respond to the lack of measured soil property information. Bouma (1989) described the term pedotransfer function as ‘translating data we have into what we need’. Pedotransfer functions are predictive functions which relate more easily measurable soil data such as soil texture (sand, silt and clay content), bulk density (BD), organic matter (OM) or organic carbon (OC) content and/or other data routinely measured or registered in soil surveys with hydraulic parameters such as the soil water retention curve (SWRC) (Bouma and van Lanen, 1987; Bouma, 1989; van den Berg et al., 1997). The most readily available data come from soil survey reports and soil databases.

A thorough review on the use and the development of hydraulic PTFs was provided by Wösten et al. (2001). Later, Shein and Arkhangel'skaya (2006) analysed the potential, state of the art, and outlooks of using PTFs in soil science. McBratney et al. (2002) reported that the estimation of soil water retention constitutes the most comprehensive research topic in development of PTFs. This may be due to particular efforts, time and cost of measurements of this hydraulic property, the need to obtain information on this property for large-scale studies, coupled with availability of large databases containing information on water retention. However, Schaap (2005) wrote that ‘with the exception of a few studies, hydraulic data and corresponding indirect methods about tropical soils are a virtual *terra incognita*’. This situation has not changed that much today. Minasny and Hartemink (2011) indicated that fewer efforts are devoted in prediction of properties of soils of the tropics where the need for accurate and up-to-date soil property information is even more urgent. They published a review paper on PTFs for predicting physical and chemical properties of soils in the tropics. First, the authors discussed the guiding principles of prediction and the type of predictors, followed by a discussion on PTFs for soil physical and chemical properties followed by infrared spectroscopy, proximal sensing and remote sensing. Several authors evaluated the prediction performance of various ‘tropical’ as well as ‘temperate’ water retention PTFs on their local soil datasets (e.g., van den Berg et al., 1997; Medina et al., 2002; Tomasella and Hodnett, 2004; Patil and Rajput, 2009; Nebel et al., 2010). Actually, the last review paper exclusively dedicated to published PTFs that predict water retention of ‘tropical’ soils was published by Tomasella and Hodnett (2004). Therefore, there is a need to provide an up-to-date repository of past and recent published articles as well as proceeding papers dealing with water retention PTFs for soils of the humid tropics. Tropical regions here are defined as regions located between 25° N and 25° S and mainly under a (sub)humid climate. Soils within the tropics but under temperate climate conditions due to altitude or in dry areas were not considered in this review. Another important contribution of this review is the categorization of published water retention PTFs based on various approaches and their application on soils of the humid tropics based on available literature data. This will allow identifying the most common and less common approaches used to estimate water retention of soils in tropical regions.

3.2 Categorization of water retention PTFs

Based on different criteria used by various authors during the last three decades, PTFs used to estimate water retention of soils can be categorized as:

- class PTFs and continuous PTFs;
- point-based PTFs, parameter-based PTFs and pseudo-continuous PTFs;
- PTFs based on a specific approach; and
- equation-based PTFs and pattern-recognition PTFs.

In this review, the terminologies ‘point-based PTFs’ and ‘parameter-based PTFs’ are preferred to the more widely used ‘point PTFs’ and ‘parametric PTFs’ in order to avoid confusion with the term ‘non-parametric PTFs’ used by some researchers (e.g., Nemes et al., 2006a,b; Nemes et al., 2008; Gharahi Ghehi et al., 2012).

3.2.1 Class and continuous PTFs

Within the PTFs used to generate soil hydraulic characteristics, Wösten et al. (1990, 1995) made a subdivision based on the amount of available information. They distinguished class PTFs and continuous PTFs.

3.2.1.1 Class PTFs

A class PTF predicts the hydraulic characteristics of a texture class (e.g., loamy sand) and is based on a preliminary grouping (Wösten et al., 1990, 1995). Therefore, class PTFs are cheaper and easier to use than continuous PTFs, because they require only identification of the texture class to which the soil belongs. However, the obtained accuracy is limited because only one average value of a hydraulic characteristic is provided for each textural class (Wösten et al., 1995). Bruand et al. (2004) distinguished six main grouping criteria for PTFs: genetic-based, horizon-based, texture groupings, grouping based on structure and BD, parent material and consecutive groupings. Several class PTFs have been developed for soils of temperate regions (e.g., Jamagne et al., 1977; Clapp and Hornberger, 1978; Rawls et al., 1982; Carsel and Parrish, 1988; Vereecken et al., 1989; Wösten et al., 1995; Wösten et al., 1999; Schaap et al. 2001; Bruand et al., 2002, 2003, 2004; Al Majou, 2008; Baker, 2008). Class PTFs are scarce for soils of the humid tropics. The class PTFs of Hodnett and Tomasella (2002) developed for the parameters of the van Genuchten (1980) equation are amongst the rare ones published for soils of tropical regions. One of the major constraints to their development is the availability of large databases of soils of the humid tropics to provide a sound statistical-based grouping. Moreover, class PTFs seem generally to be less attractive than continuous PTFs due to less flexibility and the occurrence of larger estimation errors in some cases. For instance, the results obtained by Hodnett and Tomasella (2002) showed that the use of class PTFs may lead to significant errors because of the variation within a given textural class.

3.2.1.2 Continuous PTFs

A continuous PTF is developed without grouping the data, but using the complete dataset to derive equations (Wösten et al., 1990). It estimates the hydraulic characteristics using, e.g., the actually measured percentages clay, silt and organic matter content (Wösten et al., 1995). The major part of the existing PTFs developed to date fall in this category. They will be discussed in the next sections.

3.2.2 Point-based PTFs, parameter-based PTFs and 'pseudo-continuous' PTFs

In literature, some researchers (e.g., Wösten et al., 2001; Cornelis et al., 2001; Sharma et al., 2006) make a distinction between PTFs that predict the water content at some chosen matric potentials (point-based PTFs) and PTFs that estimate the parameters of analytical expressions of the SWRC (parameter-based PTFs). They are referred to as Type 2 and Type 3 PTFs in Wösten et al. (2001). Additionally, a recently published type of PTFs that somewhat falls between the above two categories was introduced by Haghverdi et al. (2012) and referred to as 'pseudo-continuous' PTFs. Figure 3–1 provides a schematic representation of point, parameter-based and pseudo-continuous PTFs.

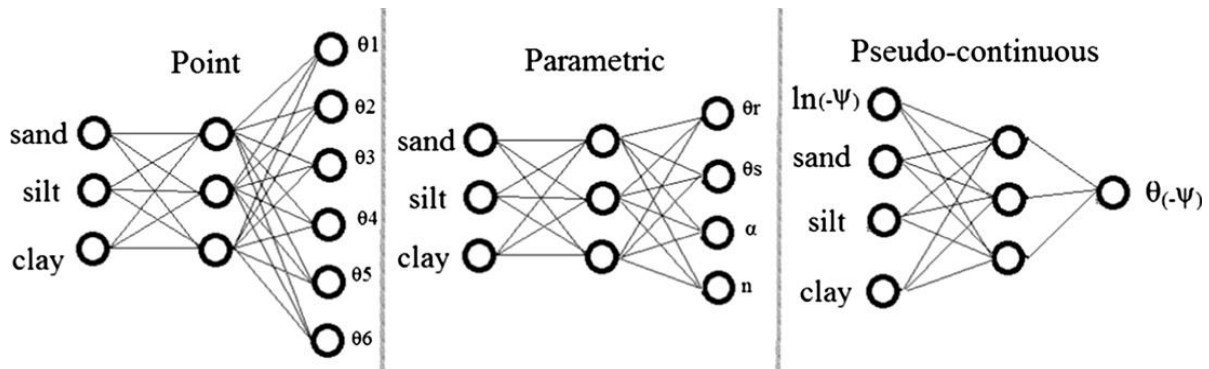


Figure 3–1. Schematic representation of point, parameter-based and pseudo-continuous pedotransfer functions (PTFs) (from Haghverdi et al., 2012).

Sand, silt and clay are sand, silt and clay percentages which are the common input predictors of PTFs. Bulk density and organic matter content can be added as optional extra predictors. $\theta_1, \theta_2, \dots, \theta_6$ are volumetric water contents which are the outputs of the point PTF when using a dataset containing six points of the water retention curve for each sample. $\theta_r, \theta_s, \alpha$ and n are the parameters of the van Genuchten (1980) equation which in turn are the outputs of the parametric PTF. $\ln(-\psi)$ is matric potential which is the extra input predictor of the pseudo-continuous PTF. $\theta(-\psi)$ is the volumetric water content at $-\psi$ matric potential which is the output predictor of the pseudo-continuous PTF. Different $-\psi$ values yield different water contents.

3.2.2.1 Point-based PTFs

The PTFs of Gupta and Larson (1979), Rawls and Brakensiek (1982), Saxton et al. (1986) are amongst the early and widely applied published point-based PTFs developed for soils in the temperate areas. Gupta and Larson (1979) used 43 different soil materials originating from ten locations in the Eastern and Central US to develop their PTFs. The twelve PTFs that estimate soil moisture content at matric potentials ranging from -4 up to -1500 kPa were developed on disturbed samples, containing mixtures of dredged sediment and productive soil in different proportions. Rawls and Brakensiek (1982) estimated water content within the same matric potential range using the same soil properties. Their data originate from 2,543 horizons from across the US. Saxton et al. (1986) developed point-based PTFs based on soils from the USDA dataset. These PTFs have been successfully applied to a wide variety of studies related to agricultural hydrology and water management, coupled with models like SPAW (Saxton and Wiley, 2006) and AquaCrop (Steduto et al., 2009). Using the currently available USDA soil database, Saxton and Rawls (2006) formulated PTFs similar to those previously reported by Saxton et al. (1986), but including more variables and having a wider application range and based on 1,722 soil sample data. In this updating process, the former equations were combined with equations of hydraulic conductivity, plus the effects of density, gravel, and salinity.

In the (sub)humid tropics, various efforts have been made to develop point-based PTFs from soil datasets specific for these regions. Most of these PTFs have been developed for application within

restricted geographical domains, and for a limited range of soil texture and soil types. Pidgeon (1972) used a dataset including a wide textural range of ferralitic soils (from loamy sands to clay) from ten sites in Uganda to derive point-based PTFs. Kaolinite was the dominant clay mineral but in two sites, the dominant clay minerals were respectively illite and montmorillonite. He estimated gravimetric water content at field capacity (FC), permanent wilting point (PWP) and available water capacity (AWC) among others. The PTF developed by Pidgeon (1972) estimates FC as equivalent to the soil moisture of a wetted plot after 48 hours of free drainage. The author also provided equations that convert this value into water content of undisturbed cores at -10 and -33 kPa. MacLean and Yager (1972) derived PTFs to predict $AWC_{10-1500 \text{ kPa}}$ based on texture, OC and soil sample depth for soils of Zambia. Simple relationships between clay content and gravimetric water content at -1500 kPa (PWP) have also been derived for ferralitic and oxic horizons in various tropical regions as reported in FAO (1974) and Soil survey staff (1975, 1990). Lal (1978, 1981) derived point-based PTFs to predict gravimetric water content at -10 kPa, -33 kPa and -1500 kPa and $AWC_{10-1500 \text{ kPa}}$ based on a dataset of soils developed on two different parent materials in Southern Nigeria. The PTF development datasets included mostly strongly weathered soils but some hydromorphic soils and high activity clay soils were also present.

Aina and Periaswamy (1985) related measured volumetric water content of undisturbed and sieved soils at -33 kPa and -1500 kPa and $AWC_{33-1500 \text{ kPa}}$ to soil texture and BD. The soil dataset comprised different Ultisols and Alfisols containing predominantly kaolinite. They constructed a PTF estimating $AWC_{33-1500 \text{ kPa}}$ for core samples from silt, clay and BD and a PTF relating $AWC_{33-1500 \text{ kPa}}$ for sieved samples to sand and BD. Arruda et al. (1987) used well drained and mainly highly weathered soils from South-East Brazil to derive gravimetric PTFs predicting water content at -33 kPa and -1500 kPa from silt and clay. Dijkerman (1988) related sand and clay content of soils from Sierra Leone to gravimetric water content at -33 kPa and -1500 kPa. The dataset used included mostly strongly weathered Ultisols and some hydromorphic soils. Bhavanarayana et al. (1986) and Rao et al. (1988) developed PTFs to predict volumetric water content at FC and PWP for Indian soils. A statistical relationship was established between clay and OC content and gravimetric water content at -1500 kPa (Soil survey staff, 1992). Bell and van Keulen (1995) derived PTFs to predict water content at PWP for four groups of Mexican soils. Van den Berg (1996) developed PTFs to predict water content at -10 and -1500 kPa and $AWC_{10-1500 \text{ kPa}}$ for strongly weathered soils in South and South-East Brazil. Van den Berg et al. (1997) used two datasets of soils originating from South America, Africa and South-East Asia. The first dataset was used to derive volumetric point-based PTFs to calculate water content at -10 kPa, -33 kPa and $AWC_{10-1500 \text{ kPa}}$ as well as to estimate the parameters of van Genuchten (1980) equation. The second dataset was used for validation purposes.

Singh (2000) derived PTFs to predict volumetric water content at saturation, at -33 kPa and -1500 kPa based on sand and clay content of soils from India. Mdemu and Mulengera (2002) developed local PTFs to predict water retention at eight different matric potentials and AWC for soils in the Morogoro region of Tanzania. Point-based PTFs were derived by Igwe et al. (2002) to predict water content of some soils of the South-eastern part of Nigeria. Tomasella et al. (2003) used a large dataset including soils from different geomorphic regions of Brazil to relate basic soil properties to volumetric water contents at different matric potentials (-6 , -10 , -33 , -100 and -1500 kPa). The point-based PTFs of Tomasella et al. (2003) use moisture equivalent as an input. It is defined as the water content remaining in a sample (fraction < 2 mm) after centrifuging at 2,400 rpm for 30 minutes, generally expressed in gravimetric units. In the PTF development process, moisture equivalent has been used by

Tomasella et al. (2003) as a predictor of water retention because it is basic information found in most Brazilian soil survey reports. Point-based PTFs were developed by Saikia and Singh (2003) to predict soil water content at FC and PWP of soils of Banha watershed of Jharkhand state in India. Likewise, Adhikary et al. (2008) developed -33 kPa and -1500 kPa PTFs based on a large collection of soil data compiled from published scientific papers, reports, and other relevant literature. Patil et al. (2009) developed -33 kPa and -1500 kPa PTFs and estimated AWC of seasonally impounded swelling-shrinking soils of Central India. Texture, BD and OC were used as predictors in most point-based PTFs developed for soils in India. Reichert et al. (2009) generated point-based PTFs to predict soil water retention at various matric potentials (-6 , -10 , -33 , -100 , -500 and -1500 kPa) based on texture, OM and BD of soils of Rio Grande do Sul in Brazil. Minasny and Hartemink (2011) developed PTFs to predict water content at -10 , -33 and -1500 kPa based on soil texture and BD. The development dataset and the validation dataset exclusively contained soils from the tropics. These soil datasets are parts of the IGBP-DIS soil database obtained from ISRIC in Wageningen (the Netherlands). Chakraborty et al. (2011) developed PTFs from a wide textural range of Indian soils for four points of the SWRC namely: -33 , -100 , -500 and -1500 kPa. Recently, Obalum and Obi (2012) proposed point-based PTFs for kaolinitic and coarse-textured ‘tropical’ soils from South-Eastern Nigeria. Santos et al. (2013) generated and validated PTFs to predict gravimetric water content at -33 kPa and -1500 kPa for different soil classes from the Central-south portion of Rio Grande do Sul State in Brazil.

3.2.2.2 Parameter-based PTFs

Parameter-based PTFs are equations that estimate the parameters of analytical expressions describing the SWRC such as the Brooks and Corey (1964), the Campbell (1974) and the widely-applied van Genuchten (1980) equations. Parameter-based PTFs generate continuous curves describing the hydraulic characteristics of soils. This is very important for modelling purposes (Tietje and Hennings, 1993; van den Berg et al., 1997; Cornelis et al., 2001) because some soil water and solute transport models require the complete SWRC as input. Furthermore, parameter-based PTFs allow the computation of hydraulic values at arbitrary pressures as indicated by Borgesen and Schaap (2005).

The first parameter-based PTFs have been developed using datasets of soils from the temperate regions. Cosby et al. (1984) and Rawls and Brakensiek (1985) developed regression equations for the model of Brooks and Corey (1964) based on soils from the US. Saxton et al. (1986) used the percentage clay and sand to calculate the parameters of a model that was derived from the SWRC model of Campbell (1974). Vereecken et al. (1989) developed widely used PTFs for the estimation of van Genuchten (1980) model parameters based on the physico-chemical characteristics (sand, clay, OC and BD) of 182 horizons of 40 different Belgian soil series. Wösten et al. (1999) predicted the parameters of the van Genuchten (1980) model using the HYPRES (Hydraulic Properties of European Soils) database including data from 4,030 horizons from all over Europe.

For tropical regions, parameter-based PTFs were developed by van den Berg et al. (1997) to predict the water retention parameters of the van Genuchten (1980) analytical equation. Tomasella and Hodnett (1998) developed PTFs to predict the parameters of Brooks and Corey (1964) equation from texture and OC using a dataset of various soils from Brazilian Amazonia. Tomasella et al. (2000) derived parameter-based PTFs for the van Genuchten (1980) model using soil information from a dataset containing 517 soil horizons from various regions in Brazil. Tomasella et al. (2000) stated that the van Genuchten (1980) analytical function is very popular in the ‘modelling community’ although it may not be the best one to properly describe the hydraulic behaviour of soils such as Oxisols. Earlier,

van den Berg et al. (1997) found that the van Genuchten (1980) equation can adequately describe moisture retention curves of soils with low-activity clays in the Southern part of Brazil. Later, Hodnett and Tomasella (2002) arrived to the same conclusion for Brazilian soils. Hodnett and Tomasella (2002) used a part of the IGBP-DIS soil database obtained from ISRIC in Wageningen (the Netherlands) to calculate the four parameters of the van Genuchten (1980) model. The authors referred to this dataset as the IGBP/T dataset which exclusively contained soils from tropical climates. Santra and Das (2008) developed parameter-based PTFs for the van Genuchten (1980) model in order to predict water retention of soils from a hilly watershed of Eastern India. Adhikary et al. (2008) did the same for the Brooks and Corey (1964) model to provide prediction based on soils from various parts of India.

From this review study, one can realize that most of the PTFs developed for soils in the (sub)humid tropics were point-based PTFs. Using validation statistics, several authors (Pachepsky et al., 1996; Tomasella et al., 2003; Dashtaki et al., 2010; Vereecken et al., 2010) noted that the point-based PTFs better predicted water retention than the parameter-based PTFs. This could be attributed to the fact that water content is controlled by different soil properties depending on the level of soil matric potentials. The point-based PTFs allow for more appropriate independent variables to describe the water content variation than do the parameter-based PTFs. This may partially explain why most of the PTFs developed for ‘tropical’ soils fall in the group of point-based PTFs. Most point-based PTFs are often limited to the prediction of water content at matric potentials generally recognized as representing FC, i.e., -10 and -33 kPa and PWP i.e. -1500 kPa. These values are typically used to calculate the water depth that should be applied by irrigation (Hansen et al., 1980) and to calculate soil water availability, which is a key element in assessing the suitability of a given region for producing a given crop (Sys et al., 1991). However, this has been perceived as a weakness by several authors (Tietje and Hennings, 1993; Cornelis et al., 2001). They argued that most of the simulation models require a continuous function rather than the discrete description provided by measurement points. Yet, van den Berg et al. (1997) and Tomasella et al. (2003) showed that point-based PTFs used to estimate soil water retention, at least in the Brazilian soils, gave more accurate results than parameter-based PTFs. Recently, Haghverdi et al. (2012) indicated that use of parameter-based PTFs presents a number of drawbacks. In some cases, the real shape of the SWRC is not similar to the shape of the chosen equation. In addition, authors like Minasny and McBratney (2002a) reported some problems in correlating the parameters of SWRC models to basic soil properties. Furthermore, parameter-based PTFs determine *a priori* which equation has to be used by the potential user, which is for most published PTFs either the van Genuchten (1980) or the Brooks and Corey (1964) closed-form equations. Vereecken et al. (2010) conducted a detailed review study on ‘temperate’ PTFs developed to estimate the parameters of the van Genuchten (1980) SWRC model.

3.2.2.3 Pseudo-continuous PTFs

Haghverdi et al. (2012) introduced pseudo-continuous PTFs, in which the natural logarithm of matric potential is considered as an input parameter, enabling the user to derive water content at any desired matric potential. Consequently, there is only one output parameter, θ , which shows the water content at the predefined matric potential i.e. different matric potential values yield different water contents. This recent approach has only been tested for soils of dry regions.

3.2.3 Specific approach-based PTFs

McBratney et al. (2002) stated that there are various ways to derive PTFs. Generally, these can be classified into two approaches:

- **semi-physical:** this approach attempts to describe a physical or chemical model relating the basic properties to the predicted properties; and
- **empirical approach:** this is the most widespread approach linking the basic soil properties (i.e. texture, pH, OC, BD) to the more difficult-to-measure soil properties by means of different numerical fitting methods.

3.2.3.1 *Semi-physical approach*

Semi-physical methods recognize the shape similarity between the shape of the particle size distribution (PSD) and water retention curves. They offer valuable conceptual insights into the physical relations between the texture distribution and the pore size distribution (POD). A drawback of these methods is that they often require very detailed PSD, making them almost as difficult to apply as direct measurements (Schaap, 2005).

Arya and Paris (1981) and Haverkamp and Parlange (1986) translated PSD data into a water retention curve by means of the capillary equation. They assumed that the network of pores in the soil is a bundle of cylindrical capillaries. Pedotransfer functions of this group require a detailed PSD (more than only clay, silt and sand content). Khlosi (2003) found that eight particle size mass fractions are sufficient to estimate the water retention curve relatively accurately. Tyler and Wheatcraft (1990) used fractal mathematics and scaled similarities to show that the empirical constant in the Arya and Paris (1981) model is equivalent to the fractal dimension of the tortuous fractal pore. The fractal dimension described by Mandelbrot (1983) is a measure of irregularity degree of the object seen in all scales (or resolutions) of observation, where the fractal structure is the one in which parts of it are similar in all. In simple words, a small piece of the object looks rather like a larger piece or the object as a whole. Therefore, the key property of fractal geometry is a degree of self-similarity across a range of spatial scales (or resolutions) of observation (Feder, 1988).

The Rieu and Sposito (1991a,b) model appears to be the first mass fractal model of the soil water retention characteristics (Millán et al., 2006). Perrier et al. (1996) developed a general model of SWRC for any soil whose POD is fractal using Mandelbrot (cumulative number versus size) distribution (Mandelbrot, 1983). Based on the previous works of Rieu and Sposito (1991a,b), Perfect et al. (1998) and Perfect (1999) used the volumetric water content definition of the prefractal Menger sponge and came up with a reduced two-parameter model of the SWRC. Perrier et al. (1999) proposed a symmetric pore-solid fractal (PSF) model. The latter is characterized by the same geometric shape in the distribution of soil pores and the distribution of soil solids; both are assumed to be power functions with the same scaling component. Based on the PSF model, Bird et al. (2000) developed a new SWRC fractal model, which includes the Tyler and Wheatcraft (1990) and Rieu and Sposito (1991a,b) models as special cases. Based on the observation that soil water retention is usually sensitive to both soil structure and texture, Millán and González-Posada (2005) assumed that two fractal regimes, each with different fractal dimension, could be present in most soils. They extended the model of de Gennes (1985) which is similar to the model of Tyler and Wheatcraft (1990), to a model with two fractal regimes. Cihan et al. (2007) introduced a general scale-variant fractal drainage model, which can be simplified into two scale-variant and scale-invariant models. In their general model, the proportion of pores which drain at a given matric potential depends on both the mass fractal dimension of the drained pore phase and the proportion of pores which drain at air-entry value. Ghanbarian-Alavijeh et al. (2010) developed a method to determine the van Genuchten (1980) model parameter m from the fractal dimension. Recently, Hwang et al. (2011) derived a symmetry and an asymmetry-based PSF model to estimate the SWRC directly from the PSD of a soil.

Studies on semi-physical models to develop hydraulic PTFs for soils in the humid tropics are scarce. Vaz et al. (2005) evaluated the performance of the Arya and Paris (1981) model applied on 104 soils from Brazil and found relatively good results. Millán and González-Posada (2005) presented a piecewise fractal approach to approximate the soil water retention data and tested their model with previously published soil datasets and two unpublished datasets corresponding to clay loam and silty clay loam soils located within a hydrographical basin in South Cuba. Andrade et al. (2008) used fractal theory to incorporate fractal dimension based on the SWRC and/or the PSD in the Brooks and Corey (1964) water retention model to estimate the available water in a soil from Brazil.

3.2.3.2 Empirical approach

The empirical approach remains the mostly used one to develop water retention PTFs in temperate, as well as in tropical regions. The most commonly used techniques for fitting or deriving PTFs are statistical regressions (Multiple Linear Regressions, MLR, and polynomials of n^{th} order). Other modern numerical and statistical methods also applied are e.g., Generalized Linear Models (GLM), General Additive Models (GAM), the Group Method of Data Handling (GMDH) and Multiple Adaptive Regression Splines (MARS). Nowadays, data mining techniques are gaining popularity in the PTF-research field with the application of non-conventional statistical methods, e.g., Artificial Neural Networks (ANNs), Classification and Regression Trees (CART), k-nearest neighbour (k-NN), Support Vector Machines (SVM), Genetic Algorithms (GA), Genetic Programming (GP).

3.2.3.2.1 Multiple linear regressions and polynomials of n^{th} order

Many of the available and well-established PTFs for the prediction of soil hydraulic properties from continuous soil properties are based on multiple linear regressions (MLR) or polynomials of n^{th} order (Vereecken and Herbst, 2004). In general, three main objectives can be distinguished when using statistical regressions to model relations between two sets of variables: prediction, model specification and parameter estimation. Multiple linear regression equations are a common statistical tool used for the prediction of the response variable y from a number of n predictor variables x_i . A multiple linear regression equation can be written as (Herbst and Diekkrüger, 2002):

$$y = a + \sum_{i=1}^n b_i x_i + \varepsilon \quad [3-1]$$

with the constant a (intercept), the regression coefficients b_i and the error ε . A nonlinear regression equation based on a second-order polynomial has the following form:

$$y = a + \sum_{i=1}^n (b_i x_i + c_i x_i^2) + \varepsilon \quad [3-2]$$

where besides from the intercept a , for every predictor variable x_i , two regression coefficients b_i and c_i have to be determined (Rawls and Brakensiek, 1985).

Gupta and Larson (1979) used MLR equations of the following form:

$$\theta_p = a(\%sand) + b(\%silt) + c(\%clay) + d(\%OM) + e(BD, Mg\ m^{-3}) \quad [3-3]$$

in order to predict the soil water content (θ_p , $m^3\ m^{-3}$) for twelve different matric potentials, where a , b , c , d and e are regression coefficients, OM is organic matter content and BD is bulk density. Intermediate values of θ could be determined by fitting one of the analytical SWRC expressions.

Rawls and Brakensiek (1982) estimated water content within the same matric potential range with the following model:

$$\theta_p = a + b(\%sand) + c(\%silt) + d(\%clay) + e(\%OM) + f(BD, Mg\ m^{-3}) + g \theta_{-33kPa} + h \theta_{-1500kPa} \quad [3-4]$$

Vereecken et al. (1989) developed widely used MLR PTFs for the estimation of van Genuchten (1980) parameters based on the physical characteristics of 182 horizons of 40 different Belgian soil series (sand, clay, OC content and BD).

Tomasella and Hodnett (1998) studied Brazilian soils and derived MLR PTFs for volumetric water contents θ at nine matric potentials:

$$\theta_p = a(\%OC) + b(\%silt) + c(\%clay) + d \quad [3-5]$$

Most of the aforementioned point- and parameter-based PTFs developed for soils of the humid tropics are MLR PTFs and some PTFs, e.g., Tomasella et al. (2000), which are polynomials of n^{th} order.

3.2.3.2.2 Extended non-linear regression

Scheinost et al. (1997) found difficulty in estimating the scaling (α) and shape (n) parameters of the van Genuchten (1980) equation using the regression approach. Realizing the overparametrization (too many adjustable parameters relative to number of data points) of the van Genuchten (1980) equation, they proposed the following approach:

- set-up the expected relationship between the parameters of the hydraulic model and soil properties; and
- insert the relationship into the model and estimate the parameters of the relationship simultaneously by fitting the extended model using nonlinear regression to all data.

This approach is referred to as extended non-linear regression (ENR) by Minasny et al. (1999). Using soils from Australia, they compared MLR and ENR approaches to develop point- and parameter-based PTFs for water retention. The authors found that ENR was the most adequate approach for parameter-based PTFs. For soils in the humid tropics, this approach has been used by Hodnett and Tomasella (2002) to develop PTFs for the parameters of the van Genuchten (1980) equation.

3.2.3.2.3 Generalized linear and additive models and multivariate adaptive regression splines

Generalized linear models (GLM) extend the linear regression models to accommodate with the non-normal response distributions (Hastie and Pregibon, 1992). The theory and applications in soil science have been reviewed by Lane (2002). For example, McKenzie and Austin (1993) used GLMs to predict soil attributes such as clay content, cation exchange capacity (CEC), pH and BD, etc. using environmental variables (geomorphic unit, local relief, etc.) as predictors. Gessler et al. (1995) used GLMs to predict the presence or absence of a bleached A2 horizon using digital terrain information.

Gessler (1996) extended this flexibility by using generalized additive models (GAM). These models attempt to characterize the non-linear effect which is not considered in GLM. This is done by allowing arbitrary smooth functions of the predictors to replace some or all the linear components of the GLM (Hastie and Tibshirani, 1990). Yet the use of GAM in the soil science literature has been minimal as pointed out by McBratney et al. (2003).

Other models such as multivariate adaptive regression splines (MARS) are used to model continuous variables (Friedman, 1991; Hastie et al., 2001). Shepherd and Walsh (2002) used MARS to develop prediction equations for a number of soil properties in Eastern Africa from NIR diffuse reflectance spectra. To the limit of our knowledge, GLM, GAM and MARS have not been used to derive hydraulic PTFs in the humid tropics.

3.2.3.2.4 Artificial neural networks

An artificial neural network (ANN) consists of many interconnected simple computational elements called nodes or neurons (Figure 3–2). Neural networks are sometimes described as ‘universal function approximators’ i.e. they can learn to approximate any continuous nonlinear function to any desired degree of accuracy (Hecht-Nielsen, 1990; Haykin, 1994). An advantage of ANN PTFs, as compared to MLR PTFs, is that they require no *a priori* concept of the relations between input and output data. During an iterative calibration procedure, the optimal relations between input and output data are found and implemented automatically. A drawback is that these relations are difficult to interpret because of the black-box nature of neural networks (Schaap and Leij, 1998).

Koekkoek and Booltink (1999) used ANNs to predict water retention at various matric potentials based on Dutch and Scottish soil datasets. Schaap et al. (1999) developed ANN PTFs to determine the parameters of the van Genuchten (1980) equation. Their ANN PTFs were based on 1,209 soil samples from the US. Schaap et al. (2001) developed the ROSETTA software, a computer program which implements four hierarchical ANN PTFs for the estimation of van Genuchten (1980) water retention parameters. This stand-alone software combines neural network analyses with the bootstrap method (Efron and Tibshirani, 1993), thus allowing the program to provide uncertainty estimates of the predicted hydraulic parameters.

Minasny and McBratney (2002b) proposed a new objective function for parameter-based ANN PTFs. The authors argued that this new method, called the ‘neuron-m method’, provides a better accuracy and less bias than the ROSETTA program. This is because the network is set up so that the predicted parameters fit the measured data, instead of training the neural network to fit the estimated parameters. Sharma et al. (2006) developed 18 models resulting from a combination of bootstrapping technique and ANN to predict moisture content at eight different matric potentials and the van Genuchten (1980) parameters for soils of the Southern Great Plains in the US.

The ANN PTFs have rarely been used to predict soil water retention of humid tropical soils. Agyare et al. (2007) developed ANN PTFs for saturated hydraulic conductivity (K_{sat}) of soils of the Volta basin in Ghana. Other studies used the ROSETTA program to derive the parameters of the van Genuchten (1980) SWRC model for swelling-shrinking soils and compared the results with locally-derived or published PTFs based essentially on MLR techniques (e.g., Patil and Rajput, 2009).

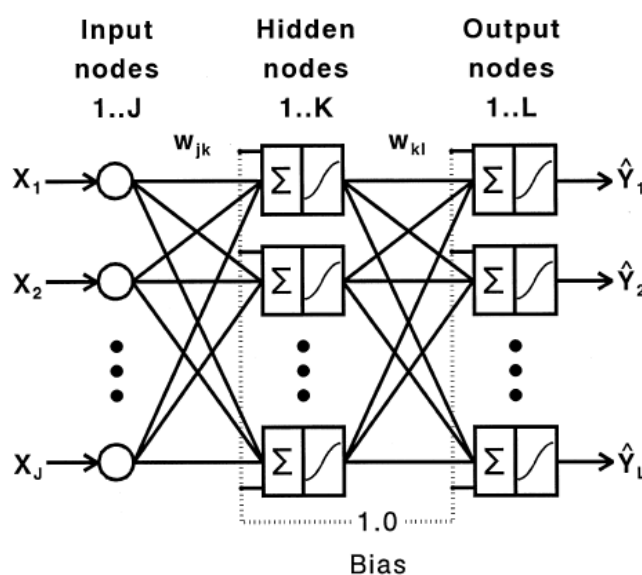


Figure 3–2. Schematic overview of a three-layer neural network (from Schaap and Leij, 1998).

3.2.3.2.5 Group method of data handling

The group method of data handling (GMDH) combines advantages of MLR techniques and ANNs (Hecht-Nielsen, 1990). The GMDH constructs a flexible neural network-type equation to relate inputs to outputs, and at the same time has a built-in algorithm to retain only essential input variables (Farlow, 1984). Some authors used this method in an attempt to improve accuracy of hydraulic PTFs of soils of temperate and tropical regions. Pachepsky et al. (1998) used GMDH to develop PTFs from texture, BD, penetration resistance, and water content at 0, -5, -10, -20, -100 and -1500 kPa in 180 soil samples from New Zealand. Nemes et al. (2005) used the GMDH technique to develop Ksat PTFs based on data originating from the US, Hungary, and the Europe-scale HYPRES database. Ungaro et al. (2005) developed hydraulic GMDH PTFs for the soils of the Pianura Padano-Veneta region in North Italy. The derived PTFs estimate Brooks and Corey (1964) water retention parameters and moisture content at -5, -10, -33 and -1500 kPa. However, they commented that their use in earth sciences is still fairly uncommon though they can provide good predictive models that successfully compete with MLR and ANN models. Tomasella et al. (2003) applied GMDH to obtain tropical point-based and parameter-based PTFs using the database of Tomasella et al. (2000), complemented with new data from a great variety of soils from Brazil. Tomasella et al. (2003) found that GMDH point-based PTFs predict water retention better than GMDH parameter-based PTFs for Brazilian soils.

3.2.3.2.6 Regression trees

A regression tree (RT) is a special type of decision tree that can predict continuous variables (McBratney et al., 2002). Regression tree modelling is an exploratory technique based on uncovering structure in data and which partitions sample data to find both the best predictors and the best grouping of samples (Clark and Pregibon, 1992). The resulting model divides data first into two groups, then into four groups and so on, providing groups as homogeneous as possible at each of the levels of partitioning. Each partitioning can be viewed as a branching, and the final fit of the model to the data looks like a tree with two branches originating in each node (Figure 3–3). Both categorical and numerical variables can be used as predictors in RT (Breiman et al., 1993).

Pachepsky and Rawls (2003) used RT in their exploration study on the potential value of structural information in the development of more accurate PTFs in modelling water transport in soils. They used data from the UNsaturated SOil hydraulic DATabase (UNSODA). Pachepsky et al. (2006) used Classification and Regression Trees (CART) to develop and discuss a PTF relating soil structure to soil water retention. This study was based on a subset of 2,149 samples from the US National Soil Characterization database. No studies were found where the CART had been applied to predict soil water content of humid tropical soils.

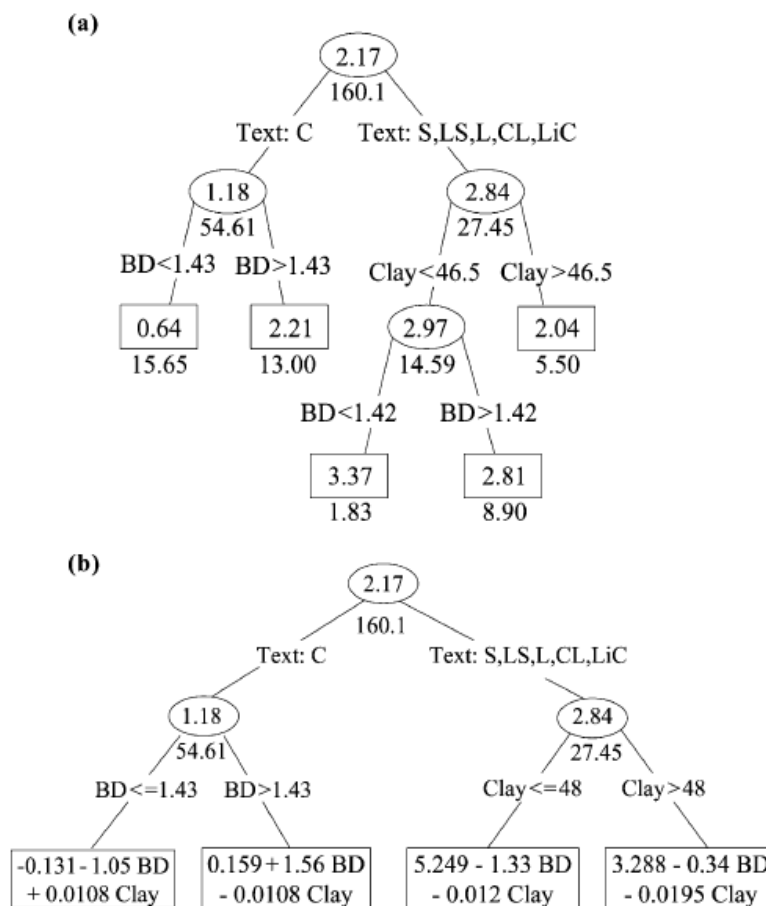


Figure 3–3. (a) A regression tree to predict saturated hydraulic conductivity (K_{sat}) from field texture grade (S = sand, LS = loamy sand, L = loam, CL = clay loam, LiC = light clay, C = clay), bulk density (BD), and clay content. Values in nodes are the predicted $\log_{10}(K_{sat})$ (in mm/day), values underneath the nodes are the standard deviation of prediction. (b) A hybrid regression tree with continuous piecewise functions (from McBratney et al., 2002).

3.2.3.2.7 k-nearest neighbour

The k-nearest neighbour (k-NN) technique is referred to as a ‘lazy learning algorithm’ that has been used for classifying sets of instances based on nearest training instances in a multi-dimensional feature space. It is said to be ‘lazy’ as it passively stores the data until the time of application. All calculations are performed ‘real-time’ i.e. only when estimations need to be generated. Once the k-NN algorithm stores a set of training instances, application of the k-NN technique means identifying and retrieving the most similar instances to the target object from that set of stored instances, based on their input attributes (Figure 3–4). More theoretical details on this similarity-based approach are given in Dasarathy (1991). The k-NN approach is considered by several authors (e.g., Buishand and Brandsma, 2001; Bannayan and Hoogenboom, 2009) as one of the most attractive pattern classification algorithms.

Nemes et al. (1999) used a k-NN variant, which they termed the ‘similarity technique’, to estimate missing soil PSD points from other existing PSD points in order to harmonize data of the European HYPRES database (Wösten et al., 1999). Jagtap et al. (2004) used a k-NN technique to estimate the drained upper limit and lower limit of plant water availability from soil water retention data measured in situ. Nemes et al. (2006a,b) developed another variant of the k-NN technique to predict soil water retention at -33 and -1500 kPa and they also performed a detailed sensitivity analysis of this technique. The newly developed k-NN algorithm proved its robustness in different scenarios. Based

on the satisfactory results yielded by their k-NN algorithm, Nemes et al. (2008) developed a user-friendly software called 'k-Nearest' with the option of estimating the uncertainty of the prediction. Elshorbagy et al. (2010a,b) identified the k-NN technique as an attractive modelling technique for hydrological applications because of its high level of flexibility. Recently, Patil et al. (2012) used the k-NN software developed by Nemes et al. (2008) to estimate water content at -33 and -1500 kPa of 157 swelling-shrinking soils in India in order to derive their AWC.

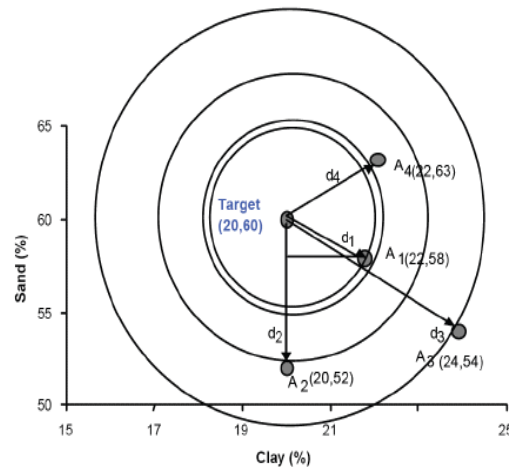


Figure 3–4. Graphical representation of k-nearest neighbour for finding the ‘best’ match for a target soil (from Jagtap et al., 2004).

3.2.3.2.8 Support vector machines

Recently, support vector machines (SVM) have gained popularity in many traditionally ANN dominated fields (Lamorski et al., 2008). They are considered as a pattern-recognition method which presents the advantage of eliminating the local minimum issue which is one of the main weaknesses of the ANN approach. Lamorski et al. (2008) used SVM to predict water retention of soils from Poland at eleven matric potentials using sand, clay and BD while Twarakavi et al. (2009) used this technique to predict the parameters of the van Genuchten (1980) equation using four different levels of the following input variables: sand, silt, clay, water content at -33 kPa and at -1500 kPa. The interested reader can refer to Vapnik (1995, 1998) and Noble (2006) for further theoretical insights. Until now, this technique has not yet been applied for soils in the humid tropics.

3.2.3.2.9 Genetic programming

Koza (1992) proposed an automatic programming technique for evolving computer programs to solve, or approximately solve, problems called genetic programming (GP). Genetic programming is a method for constructing populations of models using stochastic search methods, namely evolutionary algorithms. An important characteristic of GP is that both the variables and constants of the candidate models are optimized. Hence, compared with other regression techniques, it is not required to choose the model structure *a priori*. To date, GP has only been recently introduced in soil water related studies such as soil moisture (Makkeasorn et al., 2006), evapotranspiration (Parasuraman et al., 2007a), Ksat estimation (Parasuraman et al., 2007b), hydrological modelling (Parasuraman and Elshorbagy, 2008; Elshorbagy et al., 2010a,b; Selle and Muttill, 2011). To our knowledge, GP has not yet been used to predict water content of soils from the humid tropics.

3.2.3.2.10 Ensemble pedotransfer functions

Parasuraman et al. (2007b) stated that adopting the ensemble technique in PTF development not only assists in evaluating the uncertainty of the developed PTFs, but also helps in addressing one of the pertinent issues in any machine learning (e.g., ANNs, GP) algorithm, namely generalization. The word ‘ensemble’ is French, meaning ‘together’ or ‘at the same time’, and usually refers to a unit or group of complementary parts that contribute to a single effect. In predictive modelling, an ensemble is a set of individual models, where the component models (also known as members) are redundant in that each provides a solution to the same task, even though this solution may be obtained by different means (Baker and Ellison, 2008b). Parasuraman et al. (2006) developed ensemble ANN PTFs to estimate K_{sat} of Canadian soils whereas Baker and Ellison (2008a) developed ensemble ANN PTFs for water retention of European soils. Guber et al. (2009) selected 19 PTFs developed in different regions, among which one developed by Tomassella and Hodnett (1998) for Brazilian soils in the Amazonia region. Using the ensemble approach, they developed a new PTF to estimate soil water retention. This ensemble PTF was implemented in a programme called CalcPTF (Guber and Pachepsky, 2010). To date, there is no ensemble PTF which has been developed based exclusively on PTFs developed in tropical regions.

3.2.4 Equation-based PTFs and pattern-recognition PTFs

The PTFs described above can be categorized in equation-based PTFs and pattern-recognition PTFs. Equation-based PTFs are directly related to a mathematical model. Their formulation is based on conventional statistical procedures such as MLR, ENR, GLM, GAM, MARS, and GMDH in some extent. On the other hand, in pattern-recognition PTFs, no *a priori* model needs to be defined. They are based on pattern recognition and make use of the recently developed data-mining and machine learning techniques: ANN, RT, k-NN, SVM, GP. In Figure 3–5, a schematical representation of different types of water retention PTFs categorized according to various criteria is provided.

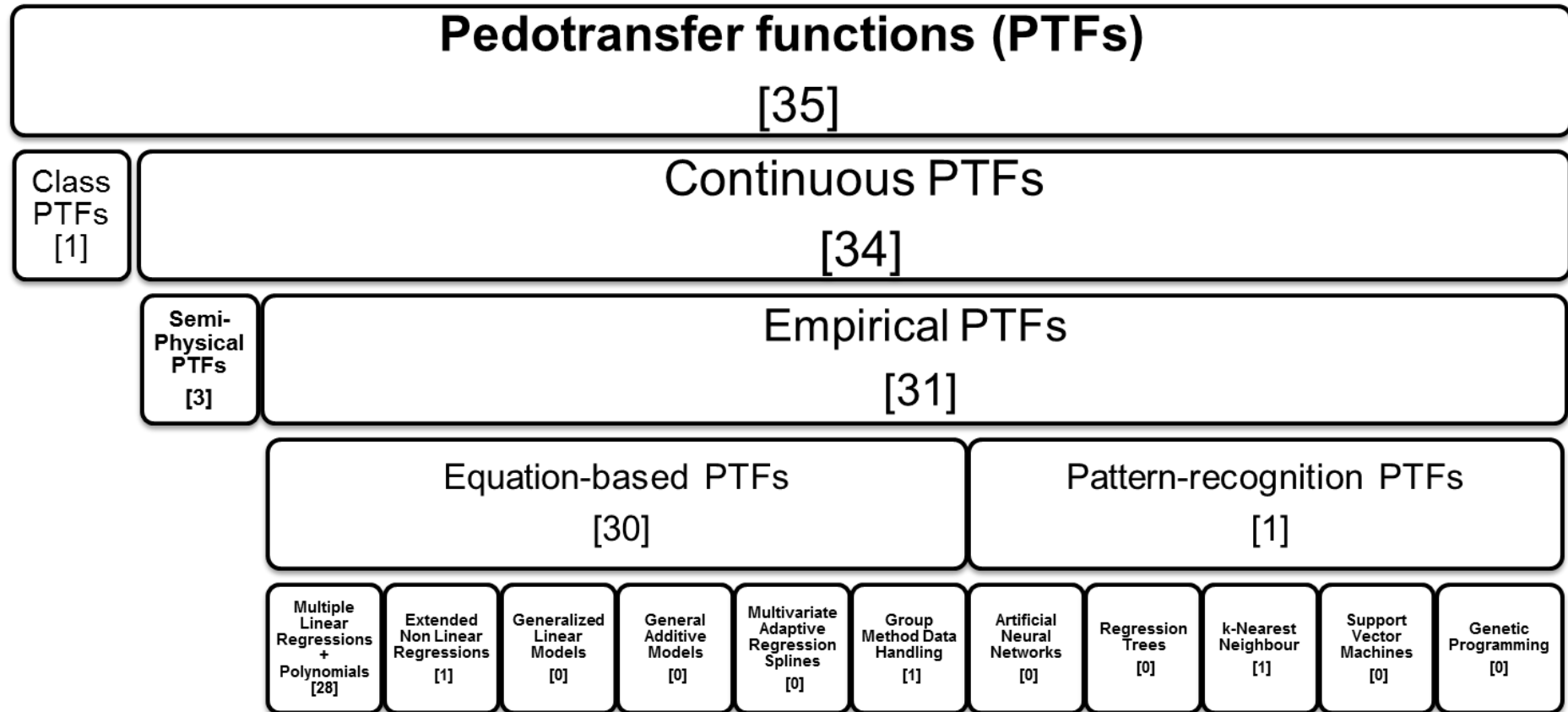


Figure 3–5. Categorization of pedotransfer functions (PTFs) to predict water retention. The number in brackets represents the number of publications in which a type of PTFs has been developed to predict water retention of soils in the humid tropics.

3.3 Summary

This review study on water retention PTFs for soils in the humid tropics reveals that:

- 97% of the PTFs based on the empirical approach are Continuous PTFs and 3% are Class PTFs;
- 91% of the PTFs are based on the empirical approach and only 9% are based on the semi-physical approach;
- 97% of the empirical PTFs were derived based on MLR and polynomial of n^{th} order techniques and 3% are based on k-NN approach.
- 84% of the continuous PTFs are point-based, 16% are parameter-based and 0% are pseudo-continuous PTFs;
- 97% of the continuous PTFs developed for soils in the humid tropics are equation-based PTFs and 3% are pattern-recognition PTFs and;
- 26% of the ‘tropical’ water retention PTFs were developed for soils in Brazil, 26% for soils in India, 11% for soils in other countries in America (US, Mexico and Cuba), 11% for soils in other countries in Africa (Nigeria, Sierra Leone, Tanzania, Uganda and Zambia).

Table 3–1 shows that the aforementioned ‘tropical’ PTFs are derived based on different scales of data collection from local to international level. Most of the soils in the development dataset were highly weathered soils dominated by low activity clay minerals such as kaolinite. They mainly belong to the WRB Reference Soil Groups of Ferralsols and related soils (Acrisols and Nitisols). However, other soils like swelling-shrinking soils (Vertisols) and hydromorphic soils were also present but to limited extent. There is a lot of variation in the size of the development dataset ranging from 13 to 685 soil samples. Most of the local tropical PTFs published in peer-reviewed journals originated from three countries located in South America (Brazil), West Africa (Nigeria) and South-East Asia (India). This shows that in other regions of the humid tropics like in Central Africa, there were no efforts in the development of local water retention PTFs.

Table 3–2 shows that different predictors were used in tropical PTFs and different output variables were considered. In Figure 3–6, one can see that PSD or texture is the dominant predictor occurring in all the ‘tropical’ PTFs’ papers (100%), followed by OC/OM (54.8%) and BD (38.7%). Therefore, these are the most common predictors in tropical PTFs compared to other soil properties found in various databases of soils from the humid tropics such as pH, CEC, DCB-Fe and DCB-Al. Only van den Berg (1996) and van den Berg et al. (1997) used DCB-Fe and DCB-Al as predictors in water retention PTFs. The results were very promising as the PTFs containing DCB-Fe and DCB-Al yielded the best results. Other soil properties such as specific surface area (van den Berg et al., 1997), moisture equivalent (Tomasella et al., 2003), micro-porosity and total porosity (Obalum and Obi, 2012) are rarely found in tropical soil databases and therefore have been rarely used as predictors in the development of ‘tropical’ PTFs. Nowadays, other predictors related to the structure of the soils (e.g., Pachepsky et al., 2006) and topography (e.g., Sharma et al., 2006) are suggested to improve the predictive ability of PTFs but have not yet been used in ‘tropical’ PTFs.

From Table 3–2, it can be seen that gravimetric or volumetric water contents at -10 , -33 , and -1500 kPa are the most selected output variables. The first two matric potentials are related to FC of ‘tropical’ soils whereas the third one is related to PWP. The PTFs of Pidgeon (1972) and van den Berg et al. (1997) predict volumetric water content at -10 kPa but not at -33 kPa like most of the aforementioned PTFs do. According to various authors (Pidgeon, 1972; Babalola, 1979; Lal, 1978; Reichardt, 1988), the water content at -33 kPa as water content at FC is considered to be too low for

‘tropical’ soils and they suggested the measure of soil moisture at -10 kPa. Other authors such as Ottoni Filho and Ottoni (2010) suggested even -6 kPa matric potential as providing a more accurate estimation of FC for Brazilian soils. Water retention at -100 kPa was also a recurrent output variable in all tropical PTFs which predict water content at several matric potentials. The critical matric potentials at which many crops undergo water stress are in the order of magnitude of -100 kPa for too dry conditions (Taylor and Ashcroft, 1972). In a similar vein, Pidgeon (1972) indicated that -100 kPa has often been considered as the upper limit of freely available water since below this matric potential the growth of many crops may be reduced.

Virtually all parameter-based tropical PTFs predict the parameters of the van Genuchten (1980) model with the exception of PTFs developed by Adhikary et al. (2008), which predict the parameters of Brooks and Corey (1964) model.

These figures confirm once more the earlier statement of Schaap (2005) that ‘with the exception of a few studies, hydraulic data and corresponding indirect methods about tropical soils are a virtual *terra incognita*’. This situation has not changed much until present. Data-mining techniques such as ANNs, k-NN and SVM which are gaining popularity in pedotransfer modelling in temperate regions have rarely been applied to develop PTFs for soils of the humid tropics. Pedotransfer functions derived from the semi-physical approach such as the pore-solid fractal (PSF) approach have not yet been applied for soils of the humid tropics. Also, more studies should be dedicated in the development of tropical PTFs based on new and/or promising predictors such as DCB-Fe.

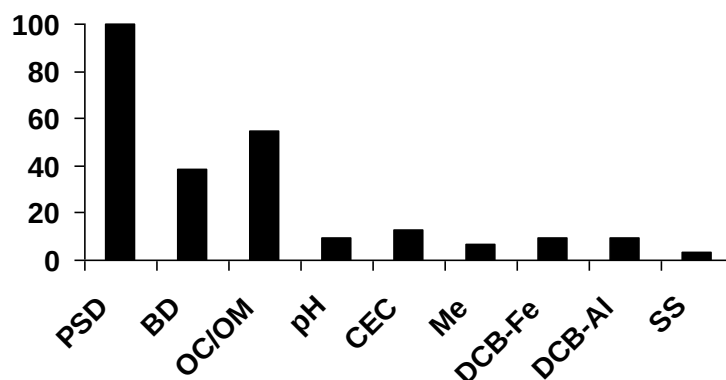


Figure 3–6. Percentage of occurrence of predictors of 'tropical' pedotransfer functions in publications. PSD is particle size distribution, BD is bulk density, OC is organic carbon, OM is organic matter, CEC is cation exchange capacity, Me is moisture equivalent, DCB-Fe is dithionite-citrate-bicarbonate extractable iron, DCB-Al is dithionite-citrate-bicarbonate extractable aluminium, SS, specific surface area.

Table 3-1. General information on point-based, parameter-based and semi-physical water retention pedotransfer functions (PTFs) developed for soils of the humid tropics.†

	Year of publication	Scale of soil data collection	Information on the soils used in the dataset	Size of the dataset	Geographical domain
Point-based PTFs					
Pidgeon	1972	Local soil survey data	Ferralitic soils dominated by Kaolinite (LAC)	39	Uganda
MacLean and Yager	1972	Local soil survey data	Oxisols, Inceptisols, Entisols and Vertisols	143	Zambia
FAO	1974	Soil survey data from various countries	Ferralic horizons (LAC)	unknown	Various countries in the humid tropics
Soil survey staff	1975; 1990	Local soil survey data	Oxic horizons (LAC)	unknown	US
Lal	1978; 1981	Local soil survey data	Mostly highly weathered soils (LAC) and some HAC and hydromorphic soils	119	Nigeria
Aina and Periaswamy	1985	Local soil survey data	Mostly Ultisols and Alfisols (LAC)	unknown	Nigeria
Bhavanarayana et al.	1986	Local soil survey data	Unknown	unknown	India
Arruda et al.	1987	Local soil survey data	Mostly highly weathered soils (LAC)	unknown	Brazil
Dijkerman	1988	Local soil survey data	Mostly Ultisols (LAC) and some hydromorphic soils	166	Sierra Leone
Rao et al.	1988	Local soil survey data	Unknown	unknown	India
Soil survey staff	1992	Local soil survey data	Oxic horizons (LAC)	unknown	US
Bell and van Keulen	1995	Local soil survey data	Alfisols, Entisols and Vertisols (LAC and HAC)	148	Mexico
Van den Berg	1996	Local soil survey data	Ferralsols and related Acrisols and Nitisols (LAC)	unknown	South- and South-East Brazil
Van den Berg et al.	1997	Soil survey data from various countries	World Oxisols and related soils (LAC)	91	Various countries in the (sub)humid tropics
Tomasella and Hodnett	1998	Local soil survey data	Mostly highly weathered soils (LAC)	196	Brazil

† LAC is low-activity clay; HAC is high-activity clay; MIX is mixed-activity clay.

Table 3–1. General information on point-based, parameter-based and semi-physical water retention pedotransfer functions (PTFs) developed for soils of the humid tropics (continued).†

	Year of publication	Scale of soil data collection	Information on the soils used in the dataset	Size of the dataset	Geographical domain
Point-based PTFs					
Singh	2000	Local soil survey data	Various soils (LAC and HAC)	256	India
Mdemu and Mulengera	2002	Local soil survey data	Nitisols, Luvisols and Ferralsols (LAC)	102	Tanzania
Igwe et al.	2002	Local soil survey data	Unknown	unknown	Nigeria
Tomasella et al.	2003	Local soil survey data	Mostly highly weathered soils (LAC)	75% of 838 (i.e. 629)	Brazil
Saikia and Singh	2003	Local soil survey data	Unknown	unknown	India
Adhikary et al.	2008	Local soil survey data	Entisols, Inceptisols, Alfisols, Vertisols, Ultisols, Oxisols and Aridisols (LAC, MIX and HAC)	600	India
Patil et al.	2009	Local soil survey data	Vertisols or Typic Haplusterts and Vertic Haplustepts (HAC)	77, 13 and 15	Madhya Pradesh region in India
Reichert et al.	2009	Local soil survey data	Unknown	249 to 685	Brazil
Minasny and Hartemink	2011	Soil survey data from various countries	Various soils (LAC, MIX, HAC)	632, 652 and 648	Various countries in the (sub)humid tropics
Chakraborty et al.	2011	Local soil survey data	Inceptisols, Alfisols, Aridisols	187	India
Obalum and Obi	2012	Local soil survey data	Typic Paleustults or Ferric Acrisols (LAC)	27, 27, 36 and 18	South-Eastern Nigeria
Santos et al.	2013	Local soil survey data	Argisols, Cambisols, Neosols and Planosols (according to EMBRAPA 2006 classification)	504 and 484	Center South portion of Rio Grande do Sul

† LAC is low-activity clay; HAC is high-activity clay; MIX is mixed-activity clay.

Table 3–1. General information on point-based, parameter-based and semi-physical water retention pedotransfer functions (PTFs) developed for soils of the humid tropics (continued).†

	Year of publication	Scale of soil data collection	Information on the soils used in the dataset	Size of the dataset	Geographical domain
Parameter-based PTFs					
Van den Berg et al.	1997	Soil survey data from various countries	World Oxisols and related soils (LAC)	91	Various countries in the (sub)humid tropics
Tomasella et al.	2000	Local soil survey data	Mostly highly weathered soils (LAC)	517	Brazil
Hodnett and Tomasella	2002	Soil survey data from various countries	World Oxisols and related soils (LAC)	492	Various countries in the (sub)humid tropics
Santra and Das	2008	Local soil survey data	unknown	100	India
Adhikary et al.	2008	Local soil survey data	Entisols, Inceptisols, Alfisols, Vertisols, Ultisols, Oxisols and Aridisols (LAC, MIX and HAC)	600	India
	Year of publication	Scale of soil data collection	Information on the soils used in the dataset	Size of the dataset	Geographical domain
Semi-physical PTFs					
Vaz et al.	2005	Local soil survey data	Quartzipsamments, Oxisols, Alfisols, Ultisols, Vertisols, Mollisols, Entisols (LAC, MIX and HAC)	104	Brazil
Millán and Gonzalez-Posada	2005	Local soil survey data	a clay loam and a silty clay loam soil	2	Cuba
Andrade et al.	2008	Local soil survey data	Neosolo Regolítico (according to EMBRAPA 1999 classification)	36	Brazil

† LAC is low-activity clay; HAC is high-activity clay; MIX is mixed-activity clay.

Table 3-2. Input and output variables in pedotransfer functions (PTFs) derived using data from soils in the humid tropics.†

	Input variables									Output variables							
	PSD	BD	OC/ OM	pH	CEC	Me	DCB- Fe	DCB- Al	SS	W ₋₁₀	W ₋₃₃	W ₋ 1500	AWC ₁₀ 1500	AWC ₃₃ 1500	>=4 WP	VG param	BC param
Point-based PTFs																	
Pidgeon (1972)	X		X							X	X	X		X			
FAO (1974)	X											X					
Soil survey staff (1975, 1990)	X											X					
Lal (1978, 1981)	X										X	X	X	?			
Aina and Periaswamy (1985)	X	X									X	X		X			
Bhavanarayana et al. (1986)	X																
Arruda et al. (1987)	X										X	X					
Dijkerman (1988)	X										X	X					
Rao et al. (1988)	X										?	X					
Soil survey staff (1992)	X		X									X					
Bell and van Keulen (1995)	X		X	X	X							X					
Van den Berg (1996)	X	X					X	X		X		X	X				
Van den Berg et al. (1997)	X	X	X		X		X	X	X	X	X	X	X			X	
Tomasella and Hodnett (1998)	X		X														

† PSD is particle size distribution, BD is bulk density, OC is organic carbon, OM is organic matter, CEC is cation exchange capacity, Me is moisture equivalent, DCB-Fe is dithionite-citrate-bicarbonate extractable iron, DCB-Al is dithionite-citrate-bicarbonate extractable aluminium, SS is specific surface area, W is gravimetric or volumetric water content, AWC is available water content, WP is water retention points, VG is van Genuchten (1980), BC is Brooks and Corey (1964), param is parameters, ? means that the information was not available to us.

Table 3–2. Input and output variables in pedotranfer functions (PTFs) derived using data from soils in the humid tropics (continued).†

	Input variables									Output variables							
	PSD	BD	OC/ OM	pH	CEC	Me	DCB- Fe	DCB- Al	SS	W ₋₁₀	W ₋₃₃	W ₋₁₅₀₀	AWC ₁₀₋₁₅₀₀	AWC ₃₃₋₁₅₀₀	>=4 WP	VG param	BC param
Point-based PTFs																	
Singh (2000)	X										X	X					
Mdemu and Mulengera (2002)	X	X	X							X	X	X			X		
Igwe et al. (2002)	X	?	?	?	?	?	?	?	?	X	X	X			X		
Tomasella et al. (2003)	X	X	X			X				X	X	X			X		
Saikia and Singh (2003)	X	?	?	?	?	?	?	?	?	?	?	X	?	?			
Adhikary et al. (2008)	X									X	X	X			X		
Patil et al. (2009)	X	X	X														
Reichert et al. (2009)	X	X	X							X	X	X			X		
Minasny and Hartemink (2011)	X	X	X							X	X	X					
Chakraborty et al. (2011)	X	X	X												X		
Obalum and Obi (2012)	X	X	X							X	X	X			X		
Santos et al. (2013)	X		X								X	X					

† PSD is particle size distribution, BD is bulk density, OC is organic carbon, OM is organic matter, CEC is cation exchange capacity, Me is moisture equivalent, DCB-Fe is dithionite-citrate-bicarbonate extractable iron, DCB-Al is dithionite-citrate-bicarbonate extractable aluminium, SS is specific surface area, W is gravimetric or volumetric water content, AWC is available water content, WP is water retention points, VG is van Genuchten (1980), BC is Brooks and Corey (1964), param is parameters, ? means that the information was not available to us.

Table 3–2. Input and output variables in pedotranfer functions (PTFs) derived using data from soils in the humid tropics (continued).†

	Input variables									Output variables							
	PSD	BD	OC/ OM	pH	CEC	Me	DCB- Fe	DCB- Al	SS	W ₋₁₀	W ₋₃₃	W ₋ 1500	AWC ₁₀₋ 1500	AWC ₃₃₋ 1500	>=4 WP	VG param	BC param
Parameter-based PTFs																	
van den Berg et al. (1997)	X		X		X		X	X								X	
Tomasella et al. (2000)	X	X	X			X										X	
Hodnett and Tomasella (2002)	X	X	X	X	X											X	
Santra and Das (2008)	X		X	X												X	
Adhikary et al. (2008)	X																X
Semi-physical PTFs																	
	PSD	BD	OC/OM	pH	CEC	Me	DCB- Fe	DCB- Al	SS	W ₋₁₀	W ₋₃₃	W ₋ 1500	AWC ₁₀₋ 1500	AWC ₃₃₋ 1500	>=4 WP	VG param	BC param
Vaz et al. (2005)	X														X		
Millán et al. (2005; 2006)	X														X		
Andrade et al. (2008)	X														X		

† PSD is particle size distribution, BD is bulk density, OC is organic carbon, OM is organic matter, CEC is cation exchange capacity, Me is moisture equivalent, DCB-Fe is dithionite-citrate-bicarbonate extractable iron, DCB-Al is dithionite-citrate-bicarbonate extractable aluminium, SS is specific surface area, W is gravimetric or volumetric water content, AWC is available water content, WP is water retention points, VG is van Genuchten (1980), BC is Brooks and Corey (1964), param is parameters, ? means that the information was not available to us.

4 Particle size distribution models for soils of the humid tropics

This chapter is based on Botula, Y.-D., Cornelis, W.M., Baert, G., Mafuka, P., Van Ranst, E., 2013. Particle size distribution models for soils of the humid tropics (article published in Journal of Soils and Sediments 13, 686-698).

4.1 Introduction

Particle size distribution (PSD) is considered as one of the most fundamental physical properties of the soil. In most soil survey reports, PSD is expressed as mass percentage of clay, silt and sand. These three fractions are used as predictors of important soil properties such as water retention curve, available water capacity, saturated hydraulic conductivity, thermal conductivity and adsorption properties of chemicals (Hillel, 1980; Wu et al., 1993; Minasny and Hartemink, 2011). Fernandez-Illescas et al. (2001) demonstrated the ecohydrological role of soil PSD in a water-limited ecosystem. Past studies suggest that changes in PSD can provide useful indications on the influences of land use, soil degradation and desertification processes on soils (e.g., Wang et al., 2008; Pachepsky et al., 1995; Su et al., 2004). In their study, Jing'an and Guojiang (1999) showed that sediment PSD is a more sensitive and more effective index of climatic and environmental changes than other geochemical indices. Shangguan et al. (2012) developed a practical 1-km-resolution dataset of soil PSD for China that is appropriate for regional land and climate modelling.

Nemes et al. (1999) mentioned that standardisation of PSD is a prerequisite to achieve compatibility of soil data among different countries using different textural classification systems. Various African countries located in the humid tropics use different textural classification schemes introduced by soil surveyors from different European countries. For example, in D.R. Congo, the INEAC system developed by Belgian soil scientists was applied. This system was essentially developed for highly weathered soils with low silt content (Jamagne, 1963). Other French-speaking countries from Central and West Africa use the French textural classification system whereas English-speaking countries use an English textural classification system. Nemes et al. (1999) remarked that Belgium, France and UK used different particle size limits in their national classification systems for soil texture. Jin et al. (2011) indicated that the latest developments in the study of PSD have focused on the application of the fractal approach. The interested reader can refer to the studies of, e.g., Pachepsky et al. (2003), Vázquez et al. (2008) and Wang et al. (2008) for more information.

Particle size distribution is typically used to predict hydraulic properties, among others. Bouma (1989) defined pedotransfer functions (PTFs) as predictive functions which relate easy-to-measure soil properties such as soil PSD, BD and OC or OM to difficult-to-measure soil hydraulic properties such as the soil water retention curve (SWRC) or hydraulic conductivity (K). Bittelli et al. (1999) indicated that a careful selection of the most appropriate PSD model is important to more precisely estimate soil hydraulic properties. Indeed, most PTFs are developed based on USDA sand, silt and clay fractions but not all texture class systems can offer that kind of information. Arya and Paris (1981), Haverkamp and Parlange (1986), Bird et al. (2000) and Hwang et al. (2011), among others, developed equations in which detailed PSD data are translated into POD and the latter is translated into a water retention curve by means of the capillary equation. These equations, referred to as physical-conceptual PTFs (group 3 PTFs in Cornelis et al., 2001), need more detailed PSD data and not only sand, silt and clay fractions. Therefore, standardisation of PSD to achieve compatibility within soil databases, and development of PTFs may require the generation of complete PSD curves and/or the prediction of unobserved points on the PSD curves. A more accurate description of texture is obtained by defining a PSD function (Bittelli et al., 1999) and several mathematical models have been proposed to accurately represent PSD of soils from various regions of the world.

However, there are only few studies which have attempted to compare different mathematical expressions of PSD (e.g., Buchan et al., 1993; Hwang et al., 2002; Hwang, 2004; Bah et al., 2009). Moreover, most of those models have been developed and tested for soils from temperate regions. Buchan et al. (1993) tested five log-normal models for 71 soils from New Zealand. Hwang et al.

(2002) extended the study of Buchan et al. (1993) by introducing two non-lognormal models, namely the Gompertz (Johnson and Kotz, 1970) and the Fredlund et al. (2000) models. They used a Korean soil dataset of 1,387 soil horizons collected from about 378 soil profiles. Hwang (2004) used the same dataset to study the influence of soil texture on the performance of nine different PSD equations. Bah et al. (2009) compared the fitting performance of seven PSD functions to sieve-pipette and laser diffraction PSD data of 55 fine-textured soil samples from New South Wales in Australia. Lima and Silva (2007) conducted a study on PSD of suspended sediments in river water. Zhao et al. (2011) worked with sediments adjacent to constructed dams in the China Loess Plateau derived from eroded parent materials. Recently, Shangguan (2012) investigated ten PSD models for the conversion of soil texture classification from ISSS (International Society of Soil Science) and Katschinski's to FAO/USDA System for soils from China.

Hwang et al. (2002) have indicated that their evaluation study on models for estimating PSD of the Korean soils should be considered as a starting point. They suggested that further studies should be conducted using other soil databases. Indeed, a soil dataset composed of soils from temperate regions is quite different from a dataset of soils from the humid tropics. Tropical soil datasets are characterised by different USDA texture classes with an underrepresentation of the silt class and the noticeable presence of soils with clay content greater 60%. On the other hand, de Condappa et al. (2008) conducted a study on the bimodal zone of the USDA soil textural triangle in which silt content is lower than both sand and clay contents. Soils associated with the bimodal zone are called 'bimodal soils'. In cases we have more observation points on PSD than sand, silt and clay fractions, bimodal soils can be defined more generally as soils for which some middle fractions are smaller than the front and the back fractions and even multimodal soils may be identified. However, in this study, the definition of bimodal soils given by de Condappa et al. (2008) has been retained. These soils are underrepresented in well-known databases like UNSODA which contain mostly soils from temperate areas. Indeed, bimodal soils are common in humid tropical regions where soils are often in the ultimate weathering stage. The concept of shape similarity between the cumulative PSD curve and the SWRC, validated for unimodal soils (e.g., Arya and Paris, 1981; Haverkamp and Parlange, 1986), implies that bimodal soils should theoretically exhibit bimodal hydraulic properties and to our knowledge, only one study was conducted to relate the PSD curves of 'tropical' soils with their SWRC using the Arya and Paris (1981) model (Vaz et al., 2005). We further only know two studies on PSD models' fitting ability for soils from the humid tropics. Silva et al. (2004) used a limited dataset of 130 soils to compare the fitting performances of 14 two- and three-parameter forms of different PSD models. Their study was based on only four measured points of the PSD curve: clay ($d < 0.002$ mm), silt ($0.002 \text{ mm} < d < 0.05$ mm), fine sand ($0.05 \text{ mm} < d < 0.2$ mm) and coarse sand ($0.2 \text{ mm} < d < 2$ mm). Bagarello et al. (2009) compared two PSD models in their study on the fitting ability of the Beerkan Estimation of Soil Transfer PSD model based on a limited dataset of 114 soils from Burundi.

Therefore, further research on the performance of models for describing PSD of soils from the humid tropics is needed. Indeed, to the limit of our knowledge, no published studies have been conducted based on a large dataset of soils from the humid tropics having a size comparable to the Korean dataset used by Hwang et al. (2002) and Hwang (2004). Moreover, the performance of PSD models for bimodal soils compared to unimodal soils has not yet been explicitly investigated in the previous studies. Therefore, the objectives of this study are (1) to evaluate the performance of various models for describing cumulative PSD of soils of the humid tropics using a comprehensive database and (2) to investigate the influence of texture and the bimodal character of 'tropical' soils on the fitting and predictive performances of these PSD models.

4.2 Materials and methods

4.2.1 Soil datasets

A large dataset of 1,412 soil samples collected from more than 300 soil profiles under different weathering stages and located in the South-western part of the D.R. Congo, referred to as Lower Congo, was used in this study. More details on the study area are given in Chapter 2. The Lower Congo dataset contains a large diversity of soils derived from various parent materials (basic, calcareous, igneous and metamorphic rocks, micaschists and sandy materials) and under different land uses (natural vegetation, forest reserve, agricultural fields, quarries among others). Eleven distinct WRB Reference Soil Groups are represented: Acrisols, Alisols, Arenosols, Cambisols, Ferralsols, Gleysols, Histosols, Lixisols, Luvisols, Podzols and Solonetz (IUSS Working Group WRB, 2006).

This dataset represents a wide range of soil textures and all the 12 USDA soil textural classes are represented but in various degrees (Figure 4–1). The sieve-pipette method (Gee and Bauder, 1986) was used for particle size analysis on air-dried fine earth samples. While the predominant textural classes were clay (30%), sandy clay loam (15%), sandy loam (13%), sand (10%), clay loam (7.5%) and loamy sand (7%), the least represented class was silt (< 1%) as shown in Table 4–1. In total, eight particle size fractions were determined: clay (< 2 μm), fine silt (2–20 μm), coarse silt (20–50 μm), very fine sand (50–100 μm), fine sand (100–250 μm), medium sand (250–500 μm), coarse sand (500–1,000 μm) and very coarse sand (1,000–2,000 μm).

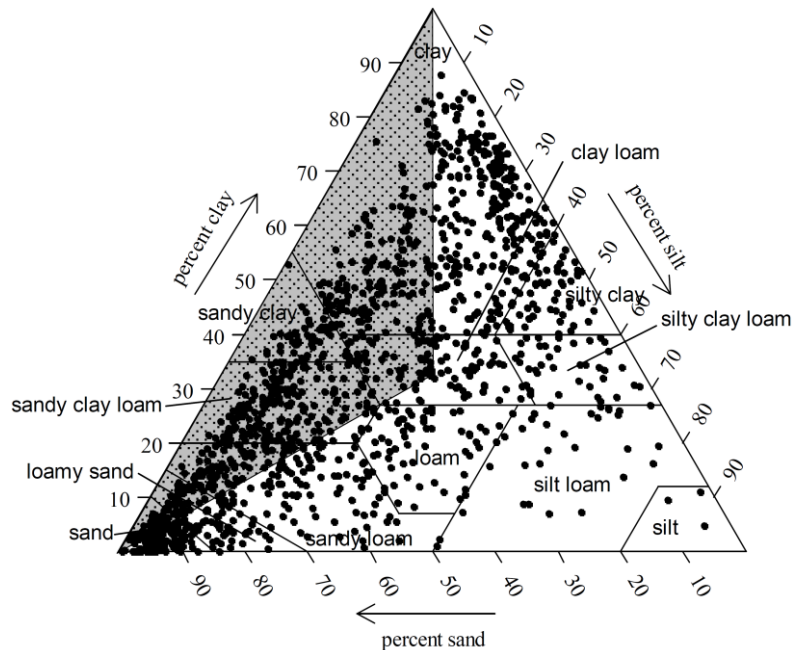


Figure 4–1. Textural composition [clay (0–2 μm), silt (2–50 μm), and sand (50–2000 μm)] of the Lower Congo dataset. The greyish area represents the bimodal zone of the USDA soil textural triangle in which silt content is lower than both sand and clay contents.

Table 4-1. Representation of the 12 USDA soil textural classes in the Lower Congo dataset by decreasing order of importance

Textural class	C	SCL	SL	S	CL	LS	L	SC	SiC	SiCL	SiL	Si	Total
Number of soils	422	213	187	147	106	97	59	52	51	40	35	3	1412
%	29.9	15.1	13.2	10.4	7.5	6.9	4.2	3.7	3.6	2.8	2.5	0.2	100

C: clay; SCL: sandy clay loam; SL: sandy loam; S: sand; CL: clay loam; LS: loamy sand; L: loam; SC: sandy clay; SiC: silty clay; SiCL: silty clay loam; SiL: silt loam; Si: silt.

In its textural composition, the Lower Congo soil dataset is strikingly similar to the ‘tropical’ subset extracted from soil datasets of the International Soils Reference and Information Centre (ISRIC) at Wageningen, the Netherlands. Referred to as the ‘IGBP/T’ dataset by Hodnett and Tomasella (2002), these authors used this subset comprising 771 tropical and subtropical soils to derive parameter-based water retention PTFs. Minasny and Hartemink (2011) also used the ‘tropical’ subset of ISRIC to develop point-based PTFs for predicting water content at -10 , -33 and $-1,500$ kPa. Unlike temperate soil datasets, such as the Soil Information System of the Netherlands (Finke, 1995) used by Nemes et al. (1999) and the Korean soil dataset used by Hwang et al. (2002) and Hwang (2004), the Lower Congo dataset and the ‘tropical’ subset of the IGBP-DIS dataset are both characterised by an important population of soils in the bimodal zone of the USDA textural triangle. This confirms that bimodal soils are common in humid tropical conditions as observed by de Condappa et al. (2008). The number of soil samples (1,412), the number of available PSD points (eight) and the large textural coverage make the Lower Congo dataset a valuable source of information to perform an evaluation study on PSD models for soils of the humid tropics.

4.2.2 Particle size distribution models

In this study, different mathematical models were used to derive continuous PSD curves (Table 4–2). Buchan (1989) proposed the simple lognormal model with two parameters (called here the LN_2p model) to describe PSD in soils. Haverkamp and Parlange (1986) used van Genuchten’s SWRC model (van Genuchten, 1980) to derive a model for the PSD. Using the relation $m=1-1/n$, they proposed a two-parameter model referred to here as the VG_2p model. Ignoring the relation $m=1-1/n$, m and n can be considered as two independent fitting parameters. This produces a three-parameter model named the VG_3p model. Fredlund et al. (2000) developed a unimodal equation to represent PSD for uniform and well-graded soils. Their five-parameter unimodal equation provided a better fit than previous two-parameter, log-normal equations for a wide variety of soils. Hwang et al. (2002) and Hwang (2004) considered one of the fitting parameters (d_m , the diameter of the minimum allowable particle size) as a constant reducing the Fredlund (2000) equation to a four-parameter model; d_m was set equal to 0.001 mm by Hwang et al. (2002) and to 0.0001 mm by Hwang (2004). In this study, this PSD model is referred to as the F_4p model with $d_m=0.0001$ mm. Working with soils from the humid tropics, Bagarello et al. (2009) have reduced the Fredlund (2000) equation to a three-parameter model by assuming the fitting parameters $d_f=0.001$ mm (a parameter related to the amount of fine particles in a soil) and $d_m=0.0001$ mm. This three-parameter form revealed excellent fitting abilities in their study and is considered as the F_3p model in the present study.

The Weibull model with three parameters was used by Assouline et al. (1998) to fit PSD curves for different soils and is referred to here as the W_3p model. A three-parameter form of the PSD model proposed by Skaggs et al. (2001) named Sk_3p was also considered in this study. Two- and four-

parameter forms of the Gompertz model (Johnson and Kotz, 1970) were used in this research and they are referred to as the G_2p and G_4p models. The G_2p equation was used by Silva et al. (2004) using a dataset of 130 soils from Brazil and the G_4p equation was used by Nemes et al. (1999) as an interpolation procedure for PSD on soils from the Netherlands and Germany. The Gompertz curve is a special case of the more general logistic curve and is described by an asymmetric closed-form equation. The G_2p and G_4p models suggest that the log mass of the soil particles follow a Gompertz distribution. As an interpolation procedure, the G_4p function considered by Nemes et al. (1999) was superior to the lognormal distribution function, which is symmetrical on the log scale. A four-parameter function reported by Jauhiainen (2004) was developed from Andersson's (1990) original theory of PSD and water retention characteristics. The author suggested that the log mass of the soil particles follow a Cauchy distribution. It is referred to here as the A_4p model. The performance of the Andersson model as well as some other PSD models was investigated and incorporated into the 'soiltexture' package of R language by Moeys and Shangguan (2010).

4.2.3 Fitting procedure

All the previously mentioned parametric functions were fitted to the observed cumulative PSD data of 1,412 soil samples using an iterative nonlinear optimization procedure. The latter finds the values of the fitting parameters giving the best fit between the model and the data (Hwang et al., 2002; Bah et al., 2009; Bagarello et al., 2009). The optimization procedure was applied using the least square curve fitting toolbox in the MATLAB R2010a environment (The MathWorks, Inc., Hill Drive Natick, MA) which is based on the 'trust-region-reflective' algorithm (Sorensen, 1982). In each case, the certainty of fitting parameter values was assured from at least three different initial parameter estimates as done by, e.g., Hwang et al. (2002) and Bah et al. (2009). The optimization process converged to very similar final values, avoiding thereby the issue of local minima.

4.2.4 Criteria for model comparison

When two or more models are constructed for the same system or purpose, comparisons between them are needed in order to select which is the best (Bellocchi et al., 2010). Many authors (e.g., Smith et al., 1997; Yang et al., 2000; Cornelis et al., 2001) advised the simultaneous use of several validation indices to perform a suitable evaluation of different models. In this study, the fitting performance of the PSD models was determined by three statistical indices, namely the adjusted coefficient of determination (R^2 -adj) [Equation 4-1], the Akaike information criterion (AIC) [Equation 4-2] (Akaike, 1973) and the relative error (E_r) [Equation 4-3]:

$$R^2_{adj} = 1 - \frac{N-1}{N-P} (1 - R^2) \quad [4-1]$$

$$\text{with } R^2 = \frac{\left(\sum_{i=1}^N (Y_{p_i} - \overline{Y_{p_i}})(Y_{m_i} - \overline{Y_{m_i}}) \right)^2}{\sum_{i=1}^N (Y_{p_i} - \overline{Y_{p_i}})^2 \sum_{i=1}^N (Y_{m_i} - \overline{Y_{m_i}})^2}$$

$$AIC = N \ln(SSE) - N \ln N + 2P \quad [4-2]$$

$$\text{with } SSE = \sum_{i=1}^N (Y_{p_i} - Y_{m_i})^2$$

$$E_r = \sqrt{\frac{\sum_{i=1}^N (Y_{p_i} - Y_{m_i})^2}{\sum_{i=1}^N (Y_{m_i})^2}} \quad [4-3]$$

where N is the number of PSD data points for soil i , P is the number of model parameters; Y_{p_i} and Y_{m_i} are predicted and measured cumulative mass fractions, respectively; $\overline{Y_{p_i}}$ and $\overline{Y_{m_i}}$ are the means of the predicted and measured cumulative mass fractions, respectively; SSE is sum of squared errors.

Table 4-2. Particle size distribution models tested in this study

Name	Model	Parameters
1 Simple lognormal (LN_2p)	$F(\ln d) = F_n\left(\frac{\ln d - a}{b}\right)$ where $F_n(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x \exp\left(-\frac{x^2}{2}\right) dx$	a, b
2 Van Genuchten type (VG_2p)	$F(d) = \left[1 + (a/d)^b\right]^{-c}$, where $c = 1 - 1/b$	a, b
3 Van Genuchten type (VG_3p)	$F(d) = \left[1 + (a/d)^b\right]^{-c}$, where b and c are independent to each other	a, b, c
4 Fredlund (F_3p)	$F(d) = \frac{1}{\ln\left[\exp(1) + \left(\frac{a}{d}\right)^b\right]^c} \left\{ 1 - \frac{\left[\ln\left(1 + \frac{d_f}{d}\right)\right]^7}{\left[\ln\left(1 + \frac{d_f}{d_m}\right)\right]^7} \right\}$	$a, b, c,$ $d_f=0.001,$ $d_m=0.0001$
5 Fredlund (F_4p)	$F(d) = \frac{1}{\ln\left[\exp(1) + \left(\frac{a}{d}\right)^b\right]^c} \left\{ 1 - \frac{\left[\ln\left(1 + \frac{d_f}{d}\right)\right]^7}{\left[\ln\left(1 + \frac{d_f}{d_m}\right)\right]^7} \right\}$	$a, b, c, d_f,$ $d_m=0.0001$
6 Weibull (W_3p)	$F(d) = c + (1 - c)\{1 - \exp(-aD^b)\}$ where $D = (d - d_{\min}) / (d_{\max} - d_{\min})$	$a, b, c,$ $d_{\min}=0.002\text{mm}$ $d_{\max}=2\text{mm}$
7 Skaggs (Sk_3p)	$F(d) = \{1 + (1/F(d_{\min}) - 1)\exp(-uD^c)\}^{-1}$ where $D = (d - d_{\min}) / d_{\min}$	$u, c, F(d_{\min}),$ $d_{\min}=0.002\text{mm}$
8 Gompertz (G_2p)	$F(d) = \exp(-\exp[-a(d - b)])$	a, b
9 Gompertz (G_4p)	$F(d) = c + e \exp(-\exp[-a(d - b)])$	a, b, c, e
10 Andersson (A_4p)	$F(d) = a + b \arctg\left[c \log\left(\frac{d}{e}\right)\right]$	a, b, c, e

$F(d)$, cumulative particle size distribution function; d , particle diameter.

Graphical representations such as clustered columns and box plots were used to provide more insights in the behaviour of the different PSD models. Soil textural classes were further grouped in three main categories according to FAO-UNESCO (1974): among the 1,412 soil samples, 41.4% were fine textured (585 soil samples), 34.2% were medium textured (483 soil samples) and 24.4% were coarse textured (344 soil samples). Soils in the dataset were also categorised in unimodal and bimodal soils according to de Condappa et al. (2008): among the 1,412 soil samples, 54.9% were unimodal (775 soil samples) and 45.1% were bimodal (637 soil samples).

The predictive ability of the best performing PSD models was investigated using a leave-one-out method. Interpolation results were shown in 1:1 plots for a visual interpretation of the predictive ability of the PSD models.

4.3 Results and discussion

4.3.1 Estimates of parameters for various PSD models

The parameters estimated for each PSD model selected in this study are shown in Table 4–3. The median values per each of the 12 USDA textural classes may serve as guides for making initial estimates in a parameter optimization process. Indeed, a good choice of initial estimates can increase the chance of the optimization routine to converge to a global minimum. It can be seen from Table 4–3 that median values of some fitting parameters can differ considerably from one textural class to the other whereas others present a lower degree of variation. A good illustration of such possible applications is given by the nonlinear least-squares optimization programme RETC (van Genuchten et al., 1991). This programme used average values for selected SWRC and K parameters provided by Carsel and Parrish (1988) as initial estimates (per USDA textural class) in the nonlinear least square optimization process.

Table 4-3. Initial estimates for the fitting parameters of the PSD models for 12 USDA soil textural classes

		C	SCL	SL	S	CL	LS	L	SC	SiC	SiCL	SiL	Si
		<i>N</i> =422	<i>N</i> =213	<i>N</i> =187	<i>N</i> =147	<i>N</i> =106	<i>N</i> =97	<i>N</i> =59	<i>N</i> =52	<i>N</i> =51	<i>N</i> =40	<i>N</i> =35	<i>N</i> =3
Simple lognormal (LN_2p)	<i>a</i>	0	0.117	0.131	0.189	0.026	0.178	0.048	0.059	0.003	0.011	0.023	0.017
	<i>b</i>	49	0.201	0.132	0.089	0.081	0.128	0.081	0.257	0.026	0.027	0.028	0.01
van Genuchten 1 (VG_2p)	<i>a</i>	0.196	0.392	0.226	0.202	0.165	0.227	0.16	0.626	0.035	0.043	0.053	0.022
	<i>b</i>	1.136	1.376	1.79	3.353	1.275	2.388	1.409	1.215	1.261	1.371	1.541	2.235
van Genuchten 2 (VG_3p)	<i>a</i>	0.336	0.642	0.359	0.283	0.289	0.338	0.278	0.957	0.051	0.059	0.061	0.034
	<i>b</i>	9.517	18.863	13.331	6.788	4.04	6.633	3.172	17.829	2.489	2.214	1.844	4.284
	<i>c</i>	0.011	0.019	0.036	0.231	0.06	0.14	0.1	0.013	0.089	0.147	0.24	0.251
Fredlund 1 (F_3p)	<i>a</i>	0.134	0.332	0.247	0.226	0.136	0.263	0.145	0.428	0.035	0.04	0.041	0.024
	<i>b</i>	1.88	3.889	3.231	4.167	1.899	3.97	1.589	4.107	1.939	1.728	1.71	3.323
	<i>c</i>	0.269	0.448	0.702	1.453	0.522	0.982	0.783	0.321	0.438	0.684	0.933	1.44
Fredlund 2 (F_4p)	<i>a</i>	0.147	0.332	0.252	0.227	0.15	0.265	0.157	0.428	0.035	0.039	0.046	0.024
	<i>b</i>	2.145	4.196	3.42	4.212	1.998	3.97	1.875	4.112	3.643	2.841	2.257	3.824
	<i>c</i>	0.216	0.431	0.676	1.424	0.471	0.971	0.687	0.313	0.197	0.402	0.749	1.257
	<i>d_f</i>	0	0	0	0	0	0	385.984	0	578.964	1211.791	444.685	103.889

C: clay; SCL: sandy clay loam; SL: sandy loam; S: sand; CL: clay loam; LS: loamy sand; L: loam; SC: sandy clay; SiC: silty clay; SiCL: silty clay loam; SiL: silt loam; Si: silt.

Table 4–3. Initial estimates for the fitting parameters of the PSD models for 12 USDA soil textural classes (continued)

		C	SCL	SL	S	CL	LS	L	SC	SiC	SiCL	SiL	Si
		N=422	N=213	N=187	N=147	N=106	N=97	N=59	N=52	N=51	N=40	N=35	N=3
Weibull	a	13.053	16.127	23.794	182.199	12.284	33.38	11.37	9.736	10.872	25.658	34.743	701.032
(W_3p)	b	0.792	1.288	1.334	2.282	0.818	1.702	0.794	1.193	0.531	0.792	0.965	1.453
	c	0.574	0.277	0.153	0.029	0.342	0.078	0.215	0.4	0.445	0.343	0.199	0.11
Skaggs	a	0.112	0.642	0.363	0.29	0.275	0.344	0.278	0.963	0.048	0.059	0.061	0.035
(Sk_3p)	c	3.813	19.252	19.321	42.168	3.361	16.788	3.169	20.882	2.602	2.214	2.465	8.586
	b	0.104	0.017	0.027	0.035	0.067	0.053	0.1	0.009	0.13	0.149	0.239	0.084
Gompertz 1	a	46.567	5.971	8.323	11.949	17.077	8.36	14.71	5.971	63.647	48.473	38.78	89.722
(G_2p)	b	0	0.043	0.076	0.145	0	0.122	0.018	0	0	0.003	0.013	0.012
Gompertz 2	c	0.402	0.142	0.042	0.025	0.009	0.048	0	0.215	0.138	0	0	0.003
(G_4p)	a	23.986	7.487	9.975	12.771	19.172	9.519	15.5	6.223	70.199	57.479	50.247	105.069
	e	0.58	0.858	0.946	0.97	0.954	0.952	0.979	0.792	0.85	0.973	0.963	0.983
	b	0	0.083	0.088	0.15	0.001	0.133	0.017	0.041	0	0.003	0.012	0.014
Andersson	a	0.784	0.666	0.597	0.53	0.68	0.562	0.623	0.729	0.718	0.65	0.589	0.533
(A_4p)	b	0.183	0.276	0.315	0.34	0.271	0.336	0.338	0.236	0.185	0.246	0.303	0.315
	c	2.261	3.371	3.688	6.919	1.968	5.091	1.823	2.921	4.178	4.024	3.194	14.099
	e	0.034	0.205	0.147	0.191	0.053	0.188	0.062	0.195	0.017	0.02	0.023	0.02

C: clay; SCL: sandy clay loam; SL: sandy loam; S: sand; CL: clay loam; LS: loamy sand; L: loam; SC: sandy clay; SiC: silty clay; SiCL: silty clay loam; SiL: silt loam; Si: silt.

4.3.2 Comparison of the fitting ability of the PSD models

In Figure 4–2, a fine and a coarse textured bimodal soil have been selected to illustrate the behaviour of the PSD models. It can be seen that some PSD models yielded a good fitting performance whereas others showed a poor fitting performance. Box plots of values for each statistical index are presented in Figure 4–3. Clear differences in model performance appear between PSD models when applied to our ‘tropical’ soil dataset. Mean R^2 -adj values are acceptable for all the models with values ranging from 0.928 to 0.993. As can be seen from Figure 4–3, F_3p, F_4p, W_3p and A_4p are the best models with the lowest mean AIC whereas the other models performed poorly with practically similar high mean AIC values. Considering average values of AIC over all the soils, the number of fitting parameters cannot always explain the difference in quality fit. Indeed, the Sk_3p model showed a lower fitting performance than other two-parameter models. On the other hand, W_3p and F_3p which have three parameters performed better than the G_4p model. However, the best PSD models were three and four-parameter models despite of the fact that AIC values impose penalties for more fitting parameters.

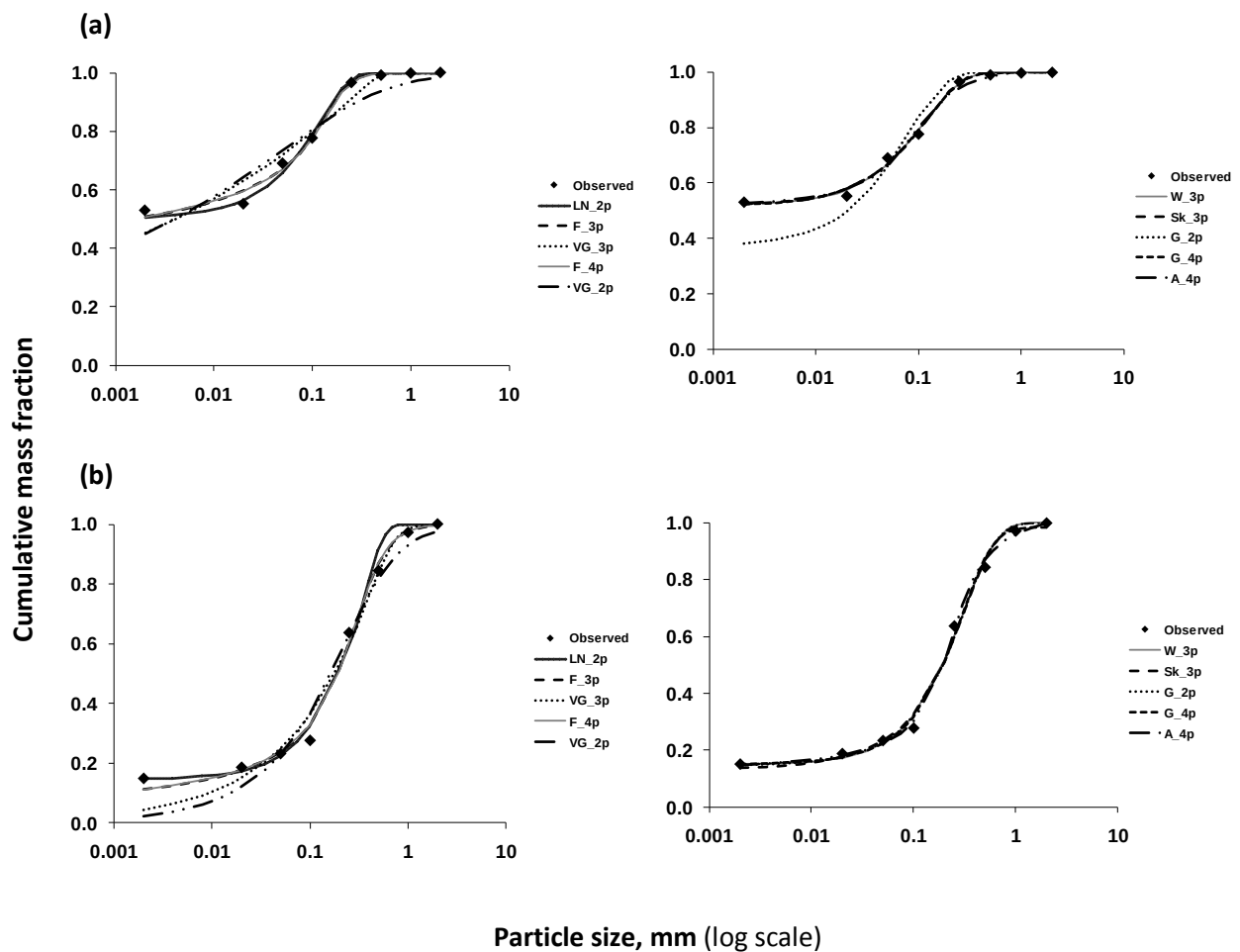


Figure 4–2. Comparative fit of ten PSD models for a bimodal (a) clay soil and (b) sandy loam soil from Lower Congo. LN_2p is Simple lognormal (2 parameters), VG_2p is Van Genuchten type (2 parameters), VG_3p is Van Genuchten type (3 parameters), F_3p is Fredlund (3 parameters), F_4p is Fredlund (4 parameters), W_3p is Weibull (3 parameters), Sk_3p is Skaggs (3 parameters), G_2p is Gompertz (2 parameters), G_4p is Gompertz (4 parameters), A_4p is Andersson (4 parameters).

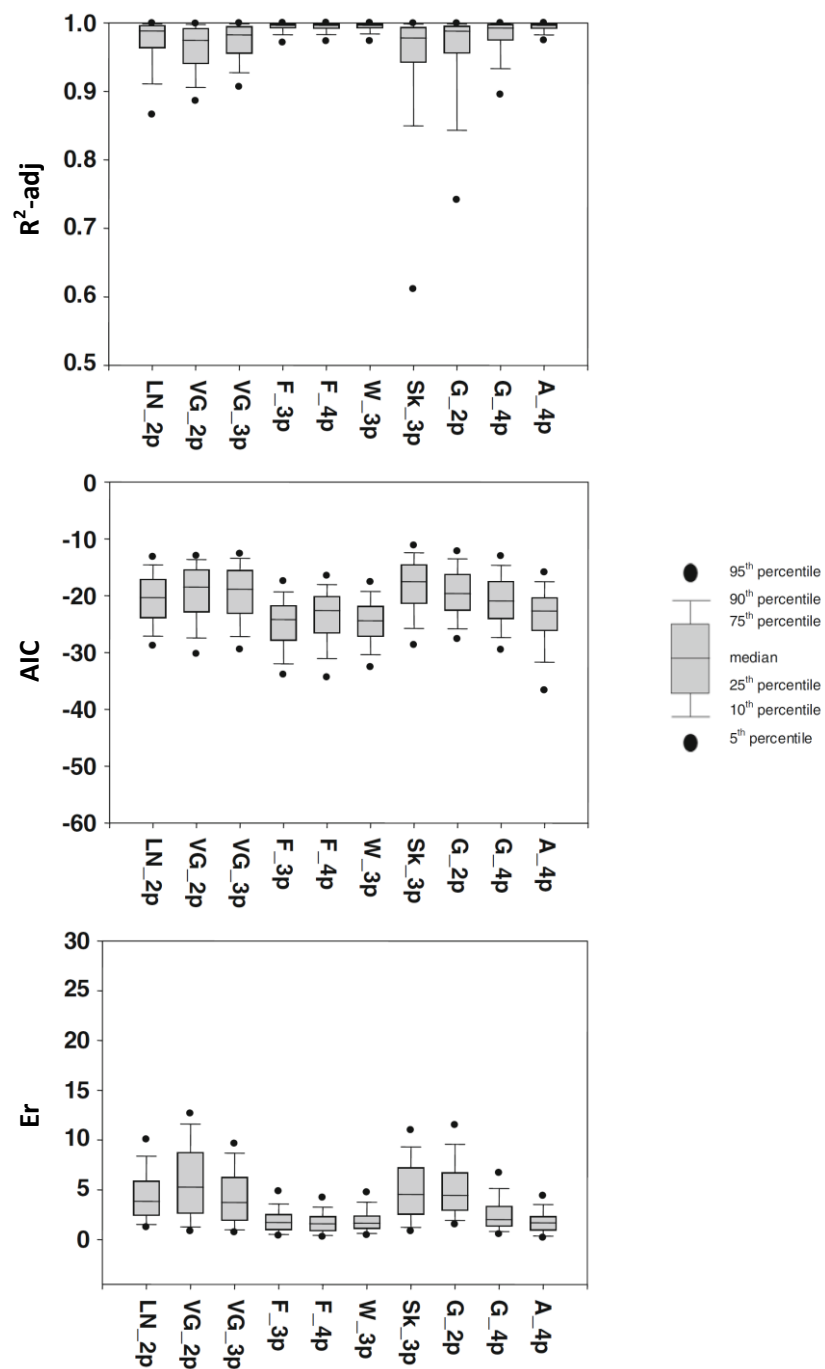


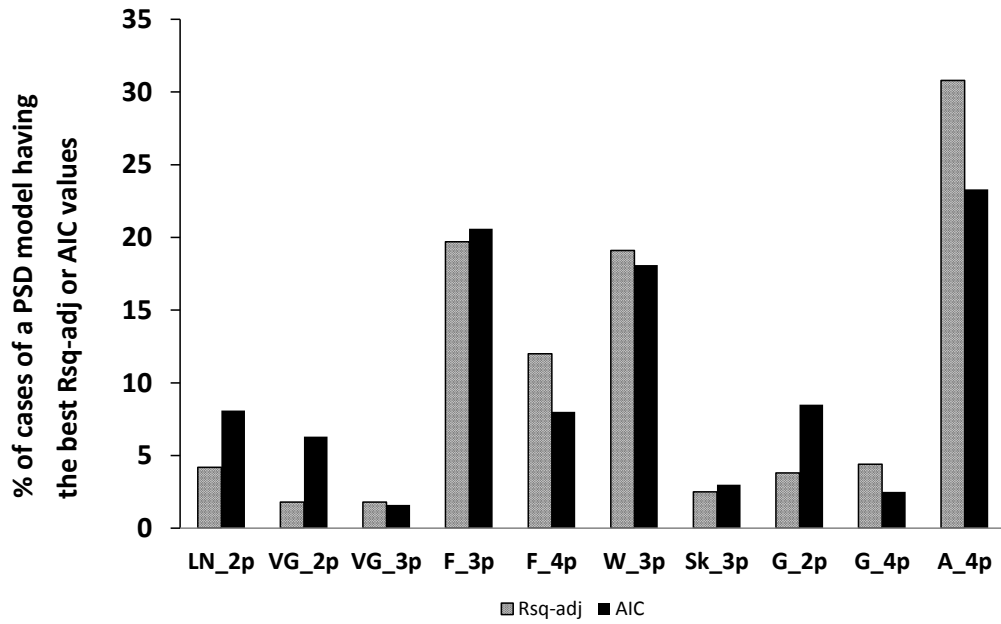
Figure 4–3. Box plots for the R^2 -adjusted (R^2 -adj), Akaike information criteria (AIC) and Relative error (E_r) percentiles of ten PSD models for 1,412 soil samples in the Lower Congo dataset.

To illustrate the importance of using different evaluation criteria for the PSD models, a graph of the percentage or number of cases where a model was the best according to a given criterion was plotted (Figure 4–4a, b). In Figure 4–4a, each model has two bars: the first bar refers to the fraction (%) of the 1,412 soil samples for which the R^2 -adj value yielded by a given model is the highest of all the R^2 -adj values yielded by the other models (indicated here as the percentage of ‘successful’ cases for R^2 -adj). The second bar shows the percentage of cases for which the AIC value yielded by a given model is the lowest of all the AIC values yielded by the other models (referred to here as the percentage of ‘successful’ cases for AIC). As depicted in Figure 4–4a, slight to more pronounced differences were found between the two criteria in selecting the best model.

Hwang (2004) worked with the four-parameter version of the Fredlund model (indicated here as F_4p) and found large discrepancies in the performance of F_4p based on R^2 and AIC values. The author observed that although the F_4p model showed an outstanding fitting performance according to R^2 criterion, its performance was not good based on the AIC criterion. This is due to the characteristic of the AIC criterion that gives more penalties to additional fitting parameters (Hwang, 2004). In this study, selection of R^2 -adj instead of R^2 can explain why the difference is less pronounced between AIC and R^2 -adj in terms of model performance. As mentioned previously, both criteria provide penalties for additional parameters. Therefore, R^2 -adj is more appropriate than R^2 criterion when comparing PSD models based on their fitting performance.

Regarding the relative error (E_r), Bagarello et al. (2009) considered that $E_r < 5\%$ indicates a satisfactory fitting performance of the PSD function. Figure 4–4b shows for each model the number of cases where $E_r < 5\%$. Vast differences exist between the models and three different levels of performance can be identified. In this study, F_4p, F_3p, W_3p, A_4p and G_4p have by far the highest number of ‘successful’ cases with more than 1,200 cases for which $E_r < 5\%$. In the second position, we find models LN_2p and VG_3p with number of ‘successful cases’ greater than 800. The remaining models (VG_2p, Sk_3p and G_2p) present a low number of successful cases. These results reflect the observations made with R^2 -adj and AIC.

(a)



(b)

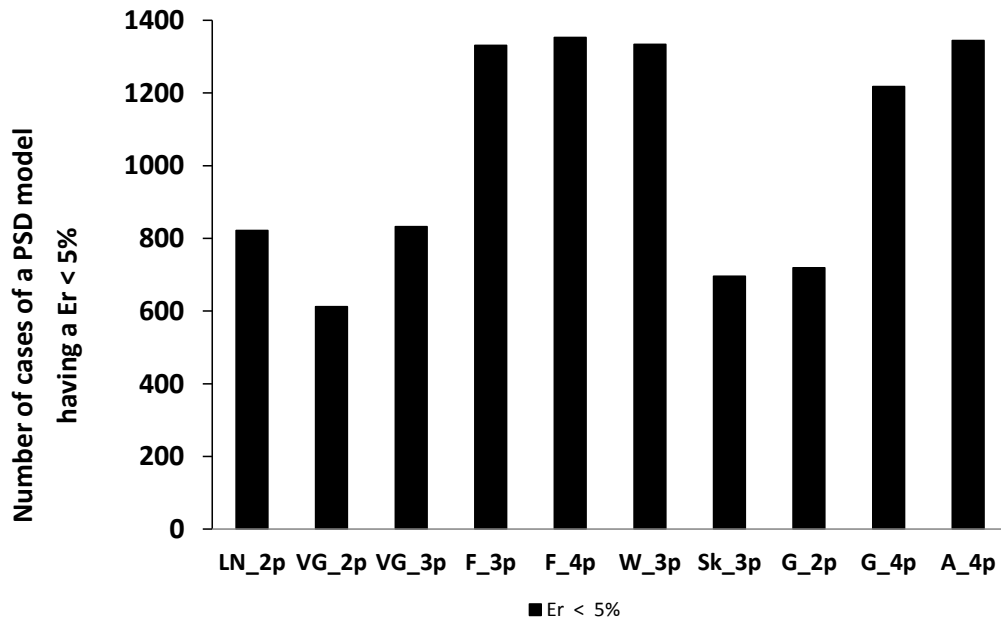


Figure 4–4. (a) Percentage of soil samples (#1,412) for which Rsq-adj or AIC are the best for a given model. (b) Number of soil samples for which the relative error (Er) is less than 5% for a given model. LN_2p is Simple lognormal (2 parameters), VG_2p is Van Genuchten type (2 parameters), VG_3p is Van Genuchten type (3 parameters), F_3p is Fredlund (3 parameters), F_4p is Fredlund (4 parameters), W_3p is Weibull (3 parameters), Sk_3p is Skaggs (3 parameters), G_2p is Gompertz (2 parameters), G_4p is Gompertz (4 parameters), A_4p is Andersson (4 parameters). (b) Number of soil samples for which the relative error (Er) is less than 5% for a given model.

4.3.3 Influence of texture on the fitting ability of the PSD models

4.3.3.1 Influence of the USDA textural class

In order to evaluate the model performance per textural class, AIC has been used as the main index by previous authors (Hwang et al., 2002; Hwang, 2004). Figure 4–5 shows the percentage of cases of a PSD model having the smallest AIC value per textural class, with subplot (a) depicting the most represented textural classes (C, CL, S, SCL, SL), and subplot (b) displaying the least represented classes (L, LS, SC, Si, SiC, SiCL, SiL). For the well-represented texture classes, the two-parameter models yielded a poor performance with less than 10% of ‘successful’ cases for CL and C. In coarse textured classes (SCL, SL and S) the performance was better for G_2p and to a limited extent for LN_2p. The three-parameter models yielded contrasting performances. The F_3p and W_3p models showed a similar performance in all the well-represented texture classes except the S class where the F_3p model showed a remarkable performance compared to other three-parameter models. In contrast, VG_3p and Sk_3p models showed very poor fitting performance (only less than 5% of ‘successful’ cases) for all the most represented classes in our ‘tropical’ soil dataset. The four-parameter models also showed contrasting performance for the C, CL, S, SCL, SL texture classes. For the finest texture class (C), A_4p is the best model with more than 40% of ‘successful’ cases whereas the rest of the models did not reach 20% of ‘successful’ cases. On the other hand, the F_4p showed less than 10% of successful cases in the most frequent textural classes in the dataset as F_3p, W_3p and A_4p scored better than F_4p.

The above results show once more that a PSD model with a higher number of parameters (e.g., G_4p or F_4p) does not necessarily perform better than one with fewer parameters (e.g., G_2p and F_3p). For instance, leaving the parameter d_f free in the four-parameter form of the Fredlund model was found to decrease its fitting performance. It was also observed that parameter d_f is particularly unstable during the fitting procedure with optimized values that can differ by several orders of magnitude (Table 4–3). This indicates that the choice of fixing the parameter d_f in the fitting process as done previously by Silva et al. (2004) and Bagarello et al. (2009) for soils in the humid tropics is justified and is therefore advised. Results displayed in Figure 4–5 demonstrate that some models are suitable for some textural classes, but are not recommended for others.

For the most frequent textural classes in our ‘tropical’ soil dataset, the F_3p and the A_4p models are the best closely followed by the W_3p model. While the F_3p and the W_3p models performed better than the A_4p model for S, SCL, CL, SL textural classes, the A_4p model showed an outstanding fitting performance for the C class. This contributed largely to the high ranking of this model when the whole Lower Congo dataset is considered. For the least frequent classes, the levels of performance seem to be more variable between the PSD models selected in this study. However no definite conclusions can be drawn due to the limited number of soils in these classes. More investigations are therefore needed with other datasets where they can be much more represented to get more reliable conclusions. Globally, the results suggest that the A_4p, the F_3p and the W_3p models satisfactorily describe the cumulative PSD of our ‘tropical’ soils.

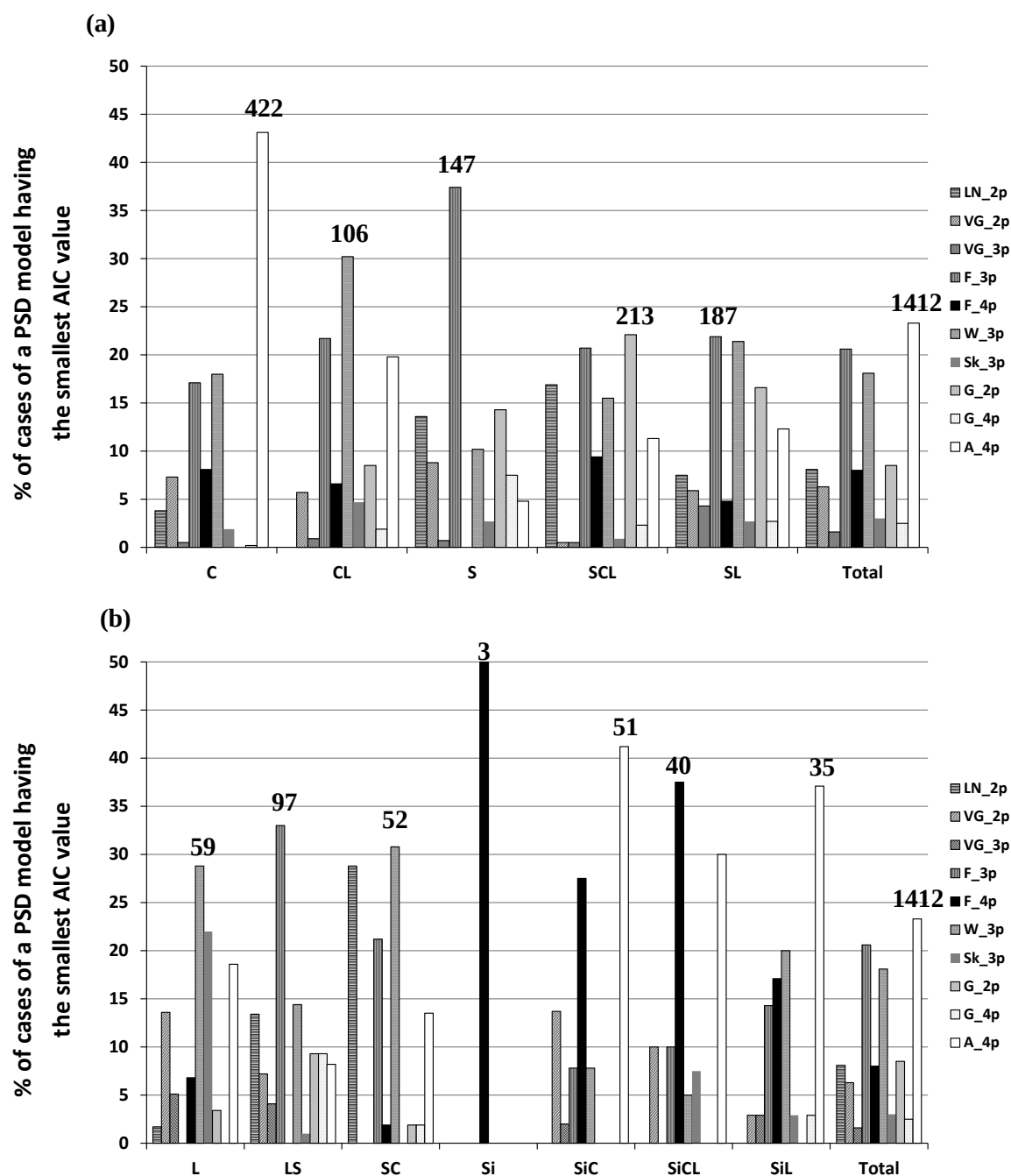


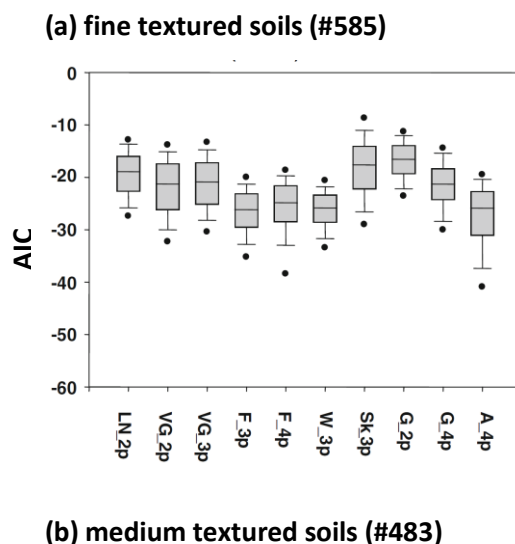
Figure 4–5. Percentage of cases for each particle size distribution (PSD) model that yielded the smallest AIC values out of a total number of soil samples for each USDA textural class, with (a) the most represented textural classes in the Lower Congo dataset and (b) the least represented textural classes in the Lower Congo dataset. The number on top of each column indicates the number of soil samples for each soil texture. C is clay, CL is clay loam, S is sand, SCL is sandy clay loam, SL is sandy loam, L is loam, LS is loamy sand, SC is sandy clay, Si is silt, SiC is silty clay, SiCL is silty clay loam, SiL is silt loam.

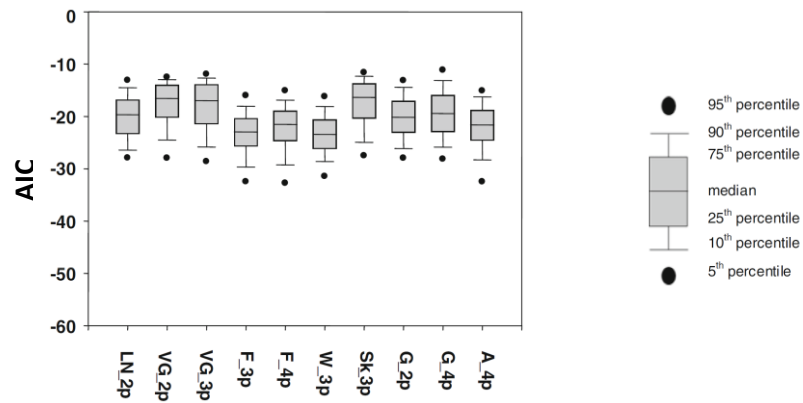
4.3.3.2 Influence of the broad textural groups

Figure 4–6 shows box plots for AIC that is used to compare the performance of the PSD models in the three broad textural groups (coarse, medium, fine). Differences in performance between the models are observed from fine- to coarse textured soils. LN_2p presents a good fitting ability for coarse-textured soils compared to fine- and medium-textured soils. In their study, Hwang and Powers (2003) demonstrated the suitability of the LN_2p model for sandy soils. In this study, F_3p, F_4p and W_3p yielded similar good performances in all three groups. However, their performance decreases from coarse texture to fine texture indicating that these models are more suitable for coarse and medium textured soils. On the other hand, the A_4p model performs better for fine-textured soils. Additionally, VG_2p, VG_3p and Sk_3p models presented similar poor fitting abilities whereas G_2p and G_4p models performed relatively better for medium and coarse textured soils. The difference in performance between the least-performing models was much more pronounced for the fine textured soils. The fitting performance of the LN_2p, G_2p and G_4p models increases from fine to coarse textured soils as opposed to the VG_2p, VG_3p models. The Sk_3p model yielded poor performance in the three broad textural groups. The G_4p model is better than the G_2p model for fine textured soils. Eventually, the F_3p and W_3p are found to be the models which describe best the PSD from fine to coarse soils in our ‘tropical’ soil dataset.

4.3.3.3 Influence of the bimodal character of ‘tropical’ soils

Figure 4–7 shows that there are differences in fitting performance of PSD models when they are applied on bimodal and on unimodal soils. At first sight, one can see that there is more variation in AIC values among unimodal soils than bimodal soils for practically all PSD models. The fitting performance of LN_2p was found to be better for bimodal soils than for unimodal soils. In contrast, VG_2p, VG_3p and Sk_3p showed lower performance for bimodal soils compared to unimodal soils. The G_2p and G_4p models presented similar fitting performance for bimodal as well as for unimodal soils. The F_3p, F_4p, W_3p and A_4p models showed the best fitting performance for both groups.





(c) coarse textured soils (#344)

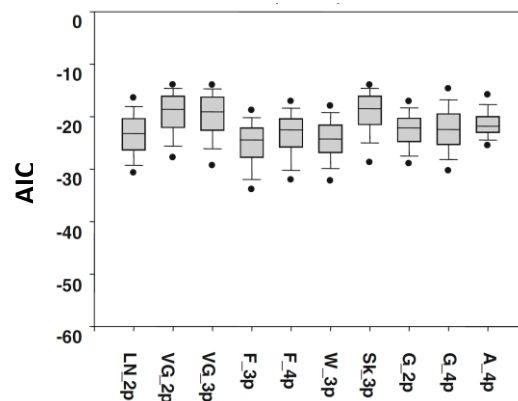
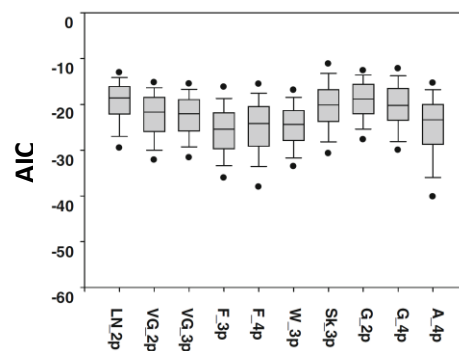


Figure 4–6. Box plots for the Akaike information criteria (AIC) percentiles of ten PSD models for 585 fine-, 483 medium-, and 344 coarse textured soil samples in the Lower Congo dataset.

(a) Unimodal soils (#775)



(b) Bimodal soils (#637)

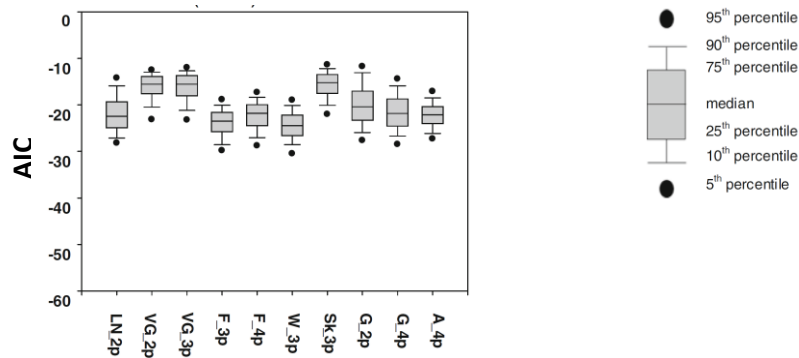


Figure 4–7. Box plots for the Akaike information criteria (AIC) percentiles of ten PSD models for 775 unimodal and 637 bimodal soil samples in the Lower Congo dataset.

4.3.4 Evaluation of prediction ability of the best fitting PSD models

The best fitting PSD models according to the first three criteria (USDA textural class, broad textural group and bimodal character), i.e. F_3p, F_4p, W_3p and A_4p, were tested for their ability to predict unknown points of the PSD curve using the leave-one-out method. As depicted in Figure 4–8, the results showed that all the four models can estimate the fine silt (2–20 μm) content of the soils with similar accuracy. However, the F_3p and F_4p models are more suited to estimate the coarse silt (20–50 μm) content than the W_3p model (slight overprediction) and the A_4p model (slight underprediction). In the sand fraction, the prediction performance of the four models can vary widely from one unknown point to the other. The F_3p, F_4p and W_3p models slightly underestimated the very fine sand (50–100 μm) content of the soils whereas the A_4p model tended to slightly overestimate it. The worst prediction performance of the aforementioned PSD models was for the fine sand (100–250 μm) content. The F_3p, F_4p and W_3p models tend to clearly underestimate the fine sand content of the soils whereas the A_4p model tends to overestimate it. The lower prediction performance of the PSD models for the fine sand content can be explained by the fact that it represents a kind of boundary between the fine part and the coarse part of the soil. This means that this fraction of the PSD of soils in the humid tropics should be preferably measured in the lab rather than interpolated using PSD models. On the other hand, the prediction of the medium and coarse sand contents of the soils was the best for the W_3p model followed by the A_4p model. The three- and four parameter forms of the Fredlund model seem to be less suitable to predict the coarsest fractions of our ‘tropical’ soils. In studies related to water retention properties of some media, Handreck (1983) observed that the soil particles in the fractions between 100 and 500 μm (i.e. from very fine sand to medium sand) controlled water release properties in horticultural growing media. Moreover, Puckett et al. (1985) used fine sand as a separate hydraulic PTF input along with the total sand as mentioned by Nemes and Rawls (2004). Based on the above results, the W_3p model seems to be the most suitable model to estimate the unknown points of the PSD for soils in the humid tropics. The trend in prediction performance between bimodal soils and unimodal soils was similar for all four PSD models as can be seen in Figure 4–8. However, it is clear that the bimodal soils show less scatter along the 1:1 line compared to the unimodal soils. The same observation was made for the least-performing models (results not shown here). This can be partly explained by a better fitting performance of the PSD models on bimodal soils compared to unimodal soils as depicted previously in Figure 4–7.

In this study, other well-known PSD models like a.o. the Jaky (1944) model borrowed from the field of geotechnics, the logarithmic (Zhuang et al., 2001), exponential (Gimenez et al., 2001) and log-exponential (Kolev et al., 1996) models, and the fractal model proposed by Kravchenko and Zhang (1998) were tested and showed poor fitting performance when applied on our ‘tropical’ soil dataset.

These results (not shown here) were in agreement with previous studies performed by several authors (e.g., Hwang, 2004; Bah et al., 2009; Zhao et al., 2011).

Based on textural composition, our ‘tropical’ soil dataset is vastly different from the Korean dataset used by Hwang et al. (2002) and later by Hwang (2004). While the Korean dataset is dominated by silt loam which constitutes 24.2% of the whole dataset, our ‘tropical’ soil dataset contains only 2.5% of silt loam soils. Figure 4–9 illustrates clearly the discrepancies between our ‘tropical’ soil dataset and the Korean dataset in terms of dominant textural classes. Carsel and Parrish (1988) wrote that kaolinitic soils of the tropics have usually clay contents ranging from 60 to 90%. This has been also observed for kaolinitic soils in our study area and depicted in Figure 4–1. Hwang (2004) observed that most of their tested models showed similar goodness-of-fits for soils with more than 60% clay. However, soils with more than 60% of clay represent less than 1% of the whole Korean dataset. One possible explanation could be that in temperate regions, soils with more than 60% of clay are considered as low permeability heavy clays and are regarded as ‘non-agricultural soils’ (Carsel and Parrish, 1988). Consequently, little can be said on the fitting ability of these models when the clay content of soils is greater than 60%. In contrast, our ‘tropical’ soil dataset contains 13.3% of soils with clay content ranging from 60 to 88%. Therefore, the evaluation of the performance of the best PSD models with increase in clay content can be better observed in our study and this on a larger interval of clay content than in the Korean dataset. Large variations in the form of the PSD curves are generally observed from more clayey to more sandy soils. This is partly due to the scale and accuracy of soil textural analyses in the lab. Among the eight lab-determined PSD points, five are in the sand fraction, two in the silt fraction and only one is in the clay fraction as indicated previously. A soil with high clay content and low sand content will have a completely different shape than a soil with high sand content and low clay content. For example, high clay content raises the first point of Figure 4–2 and makes the slope of cumulative PSD less steep. As these models are used to describe mathematically the PSD curve, it is therefore necessary to investigate the influence of clay content on their fitting performance. Our ‘tropical’ soil dataset enables a close examination of the performance of the aforementioned four best PSD models for soils with clay content greater than 60%.

Figure 4–10 presents the distribution of sum of squared errors (SSE) values of the four best PSD models as a function of clay content of Lower Congo soils. Clay content seems not to influence the performance of F_3p, F_4p, W_3p and A_4p models in describing the PSD of our ‘tropical’ soils. For the least-performing models, the SSE values seem to vary in different degrees with the clay content of the soils (results not shown here). Except very few cases, all the SSE values yielded by the four PSD models were lower than 0.05. Some very few outlying SSE values found were mostly related to soils which are gap graded, i.e. which lack one or more ranges of grain size. However, the four models are flexible enough to accommodate most of the soils of our ‘tropical’ soil dataset. Therefore, the F_3p, F_4p, W_3p and A_4p models are highly recommended to describe the PSD of soils in the humid tropics.

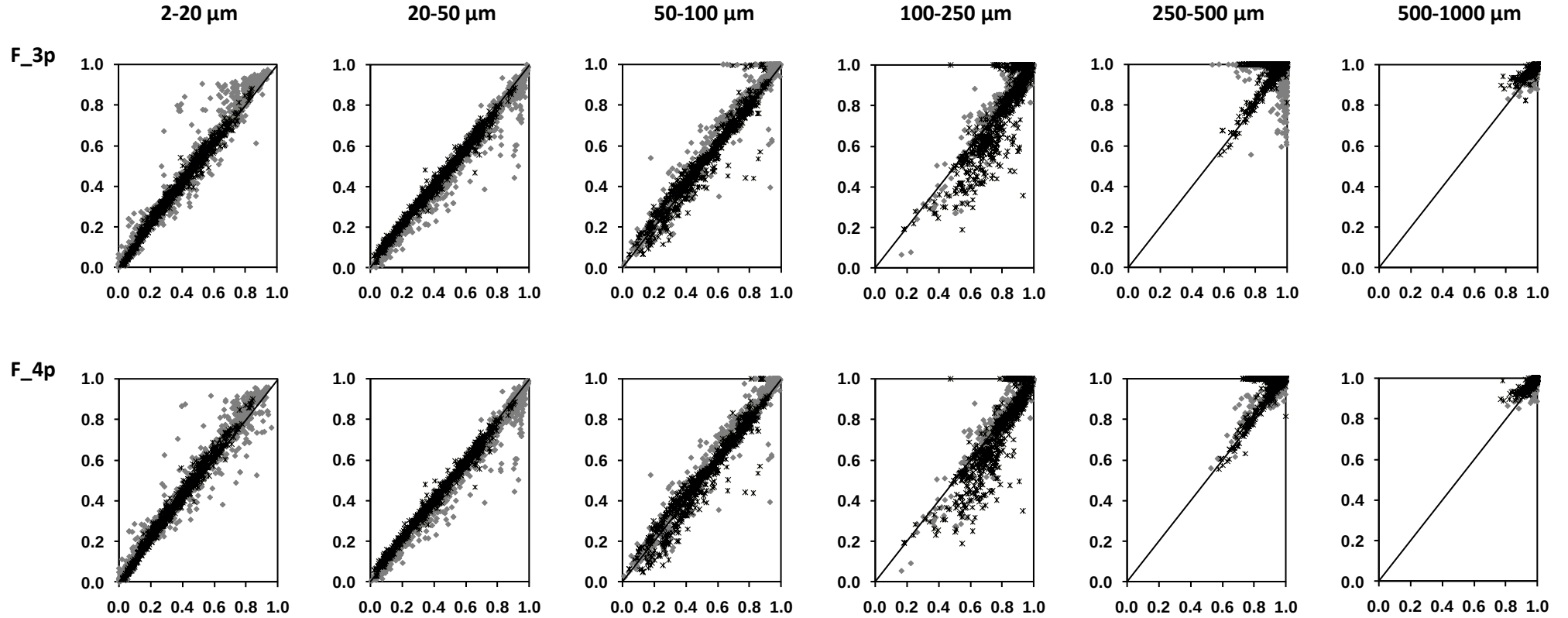


Figure 4-8. Observed particle size fractions versus particle size fractions predicted by the Fredlund-type1 (F_3p), Fredlund-type2 (Fredlund_4p), Weibull (W_3p) and Andersson (A_4p) models applied on 1,412 soil samples from Lower Congo. Grey points represent unimodal soil samples (#775) and crosses represent bimodal soil samples (#637). The figure in μm above each graph represents the USDA particle size limit for which the missing fraction needs to be predicted (interpolated) when particle size fractions within other particle size limits are known.

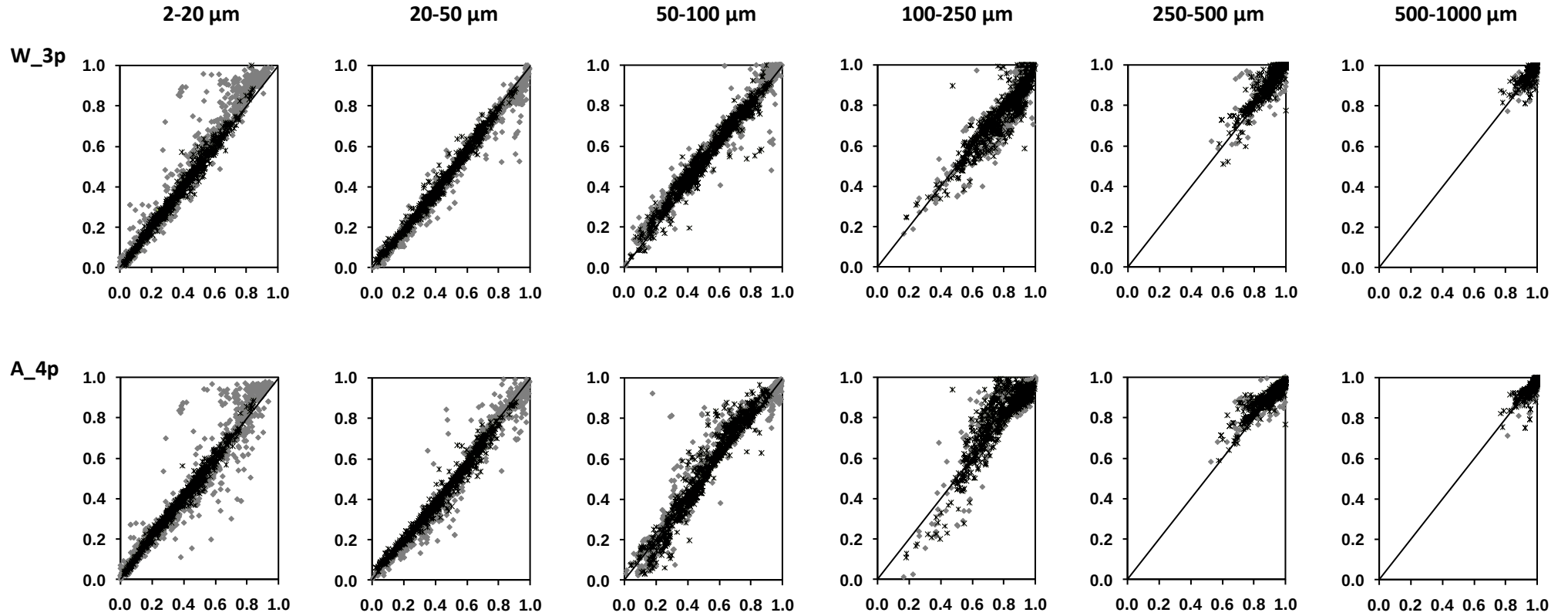


Figure 4-9. Observed particle size fractions versus particle size fractions predicted by the Fredlund-type1 (F_3p), Fredlund-type2 (Fredlund_4p), Weibull (W_3p) and Andersson (A_4p) models applied on 1,412 soil samples from Lower Congo. Grey points represent unimodal soil samples (#775) and crosses represent bimodal soil samples (#637). The figure in μm above each graph represents the USDA particle size limit for which the missing fraction needs to be predicted (interpolated) when particle size fractions within other particle size limits are known (continued).

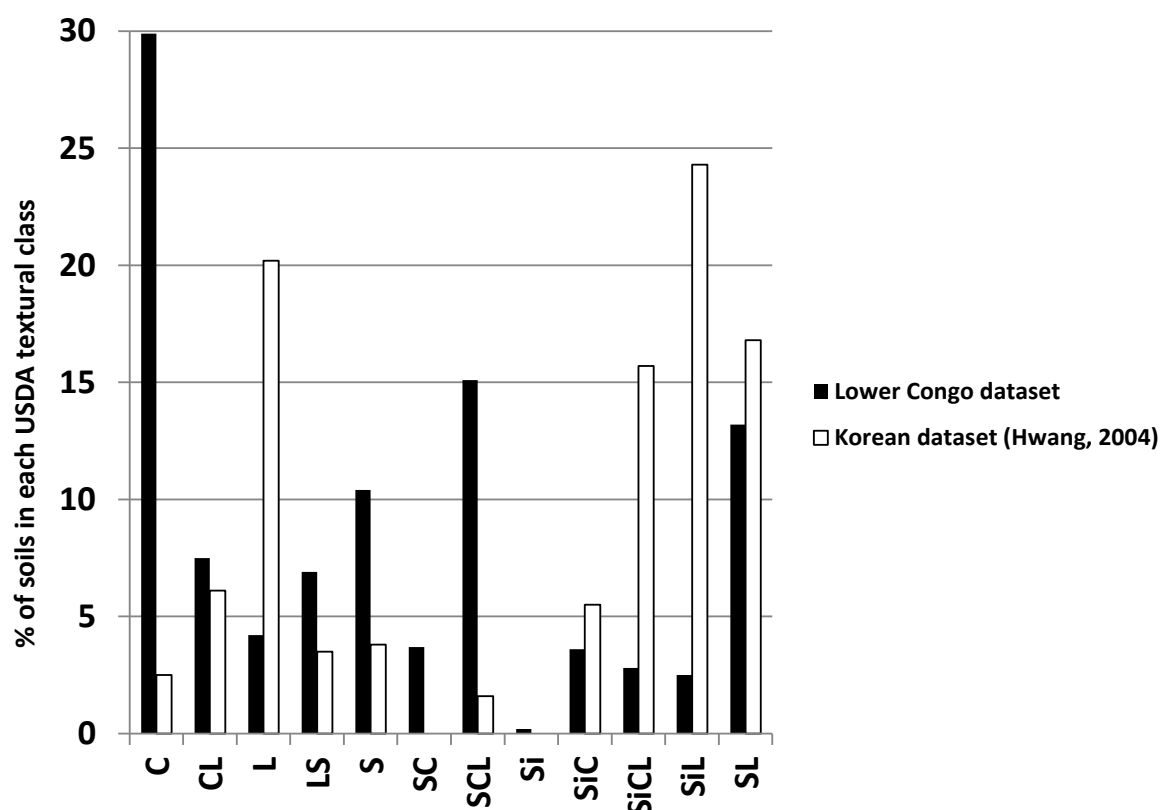


Figure 4–10. Difference in textural composition between the Lower Congo dataset (1,412 soil samples) used in the present study and the Korean dataset (1,385 soil samples) used by Hwang (2004).

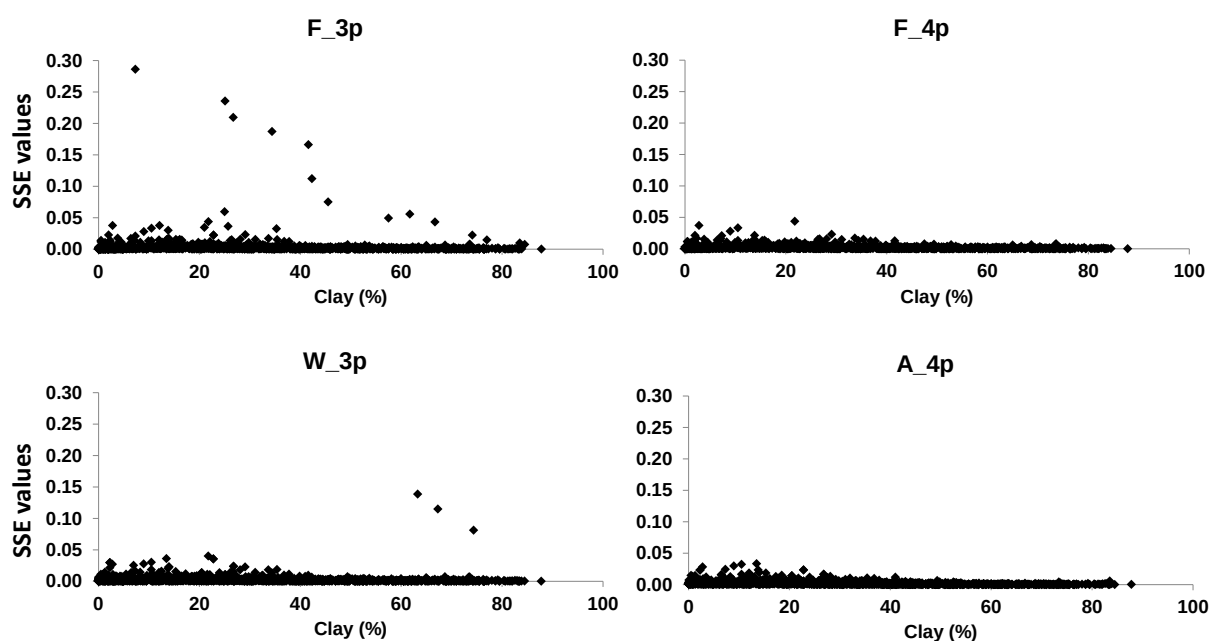


Figure 4–11. Distribution of the SSE values as a function of clay content for the Fredlund-type1 (F_3p), Fredlund-type2 (Fredlund_4p), Weibull (W_3p) and Andersson (A_4p) models for 1,412 soil samples in the Lower Congo dataset.

4.4 Conclusions

In this study, ten PSD models with one to four parameters were evaluated using a dataset of 1,412 soils from the humid tropics, collected in the Lower Congo region (D.R. Congo). Three statistical indices (R^2 -adj, AIC and E_r) were used in combination to rank the different models based on their fitting performance. Clustered columns, box plots and 1:1 plots were also used to study the influence of different categories of texture classes as well as of clay content in the performance of the models. Differences in fitting and prediction performance were found between the PSD models, and texture and the bimodal character of some soils has an influence on their respective performance. Some models like LN_2p, VG_2p, VG_3p, Sk_3p, G_2p and G_4p are generally not suitable to describe PSD of soils of the humid tropics. On the other hand, F_3p, F_4p, W_3p and A_4p models showed outstanding fitting performance. Therefore, they are highly recommended in order to obtain a better description of the PSD of soils of the humid tropics.

5 Evaluation of pedotransfer functions for predicting water retention of soils in Lower Congo (D.R. Congo)

This chapter is based on Botula, Y.D., Cornelis, W.M., Baert, G., Van Ranst, E., 2012. Evaluation of pedotransfer functions for predicting water retention of soils in the Lower Congo (D.R. Congo) (article published in *Agricultural Water Management Journal* 111, 1-10).

5.1 Introduction

Important issues like shortage of water for agriculture due to climate change, soil and groundwater pollution due to continuous application of fertilizers and pesticides, increasing costs of irrigation systems associated with the rising price of fuel put farmers under high pressure. To sustain their activities, they need to dramatically increase water use efficiency and water productivity. To help them in this task, scientists have developed models of various levels of complexity. Models like SPAW (Saxton and Willey, 2006) and AquaCrop (Steduto et al., 2009) are applied world-wide in several studies related to soil water management in agriculture. These models need data on soil hydraulic properties particularly water content at field capacity (FC) and permanent wilting point (PWP). In the literature, FC and the PWP represent two important points of the soil water retention curve (SWRC). The SWRC describes the functional relationship between the matric potential of the soil and its gravimetric or volumetric water content w or θ (Hillel, 1998; Jury and Horton, 2004) as depicted in Figure 5–1. Since the shape of the curve is related to various physical and chemical soil properties, which are unique for each soil, Cornelis et al. (2001) considered it as the fingerprint of a soil. It is generally admitted that matric potentials at -33 kPa and -1500 kPa represent FC and PWP respectively. However, direct measurements of soil water content at -33 kPa and -1500 kPa are still expensive and time-consuming especially for environmental studies at a regional scale (Vereecken, 1995). Therefore, an attractive alternative to the direct and often cumbersome measurements of these soil hydraulic properties is their estimation by so-called pedotransfer functions (PTFs). Pedotransfer functions are predictive functions which relate more easily measurable soil data such as soil texture (sand, silt and clay), bulk density (BD), organic matter (OM) (or organic carbon, OC) content and/or other data routinely measured or registered in soil surveys with hydraulic parameters such as the SWRC (Bouma and Van Lanen, 1987; Bouma, 1989; van den Berg et al., 1997).

Some water management models such as SPAW (Saxton and Willey, 2006) provide ready-to-use hydraulic PTFs to estimate water content at -33 kPa and -1500 kPa in order to respond to the potential lack of soil hydraulic data. Unfortunately, potential users working in tropical regions are generally neglecting the fact that the PTFs provided by models like SPAW were formulated based on soils from temperate areas. They blindly apply them to get estimates of water retention data for soils in the tropical regions and proceed further with the modelling process.

During the last decades, most existing and published PTFs have been developed using data from soils of temperate regions (Tomasella et al., 2000). Yet, van den Berg et al. (1997) argue that physical and chemical differences between ‘temperate’ and ‘humid tropical’ soils might be the causes of the poor performance of ‘temperate’ soils PTFs when applied to highly weathered ‘tropical’ soils. These differences originate from marked differences in mineralogical properties (Tomasella and Hodnett, 1998, 2004). Soils of the temperate regions are indicated further in the text as ‘temperate’ soils and soils of the humid tropics as ‘tropical’ soils.

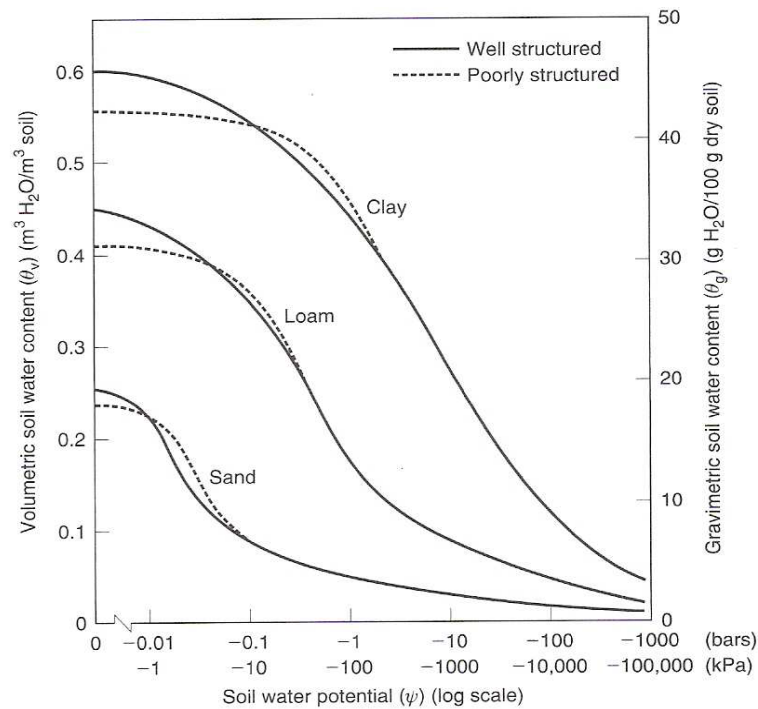


Figure 5–1. Soil water retention curves of a sand, a loam, and a clay (Source: Brady and Weil, 1999).

Based on what is indicated above, we consider that a low quality of the estimation of hydraulic properties may influence the overall quality of the outputs of the whole modelling process. Espino et al. (1996) studied the influence of ‘temperate’ PTFs of Vereecken et al. (1989) on the outcomes of the SWATRER model in the simulation of the hydraulic response of a multi-layered soil profile under natural climatic boundary conditions for a period of one year in Germany. They recommended that in further studies, many other PTFs should be included in a testing procedure prior to their application in more advanced models. To do this, a limited number of measured water retention data from the site of interest can be used to evaluate the PTFs. Most studies have evaluated PTFs that have been developed in temperate regions by using independent datasets (e.g., Tietje and Tapkenhinrichs, 1993; Kern, 1995; Imam et al., 1999; Cornelis et al., 2001; Buccigrossi et al., 2010). However, only a limited number of studies have been conducted with databases of soils from the humid tropics (e.g., Tomasella and Hodnett, 2004; Reichert et al., 2009; Nebel et al., 2010). To our knowledge, there are no studies comparing the prediction performance of both tropical and temperate water retention PTFs based on an independent database of soils from Central Africa. The evaluation dataset used in this study has the particularity to contain soils derived from various parent materials. Different weathering stages of the parent materials (recent, intermediate and ultimate) are also taken into account.

The main objective of this study is to evaluate the ability of some well-documented ‘temperate’ and ‘tropical’ PTFs to predict water retention of soils at FC and at PWP in Lower Congo. This evaluation is important to assess the necessity of improving existing PTFs and/or developing new PTFs to better estimate water retention of soils in this region. This paper will also attempt to identify some potential predictors that can contribute to the improvement of water retention prediction of soils in the humid tropics. Moreover, we consider that users of water management models in agriculture are not sufficiently aware of the effects of using ‘temperate’ PTFs on the quality of the estimation of the hydraulic properties of soils in the humid tropics. Therefore, this research can be viewed as an attempt to increase the awareness of modellers and potential users on large discrepancies in prediction performance that may occur between temperate and tropical PTFs. This study provides information on

the uncertainty (in terms of root mean squared error) when using existing PTFs to predict water retention of soils from a region in the tropics outside the one from which the PTF development database was collected. Therefore, it contributes to the general discussion about reliability of PTFs.

5.2 Materials and methods

5.2.1 The study area

The study area has been described in Chapter 2. Various locations of the sites where soil samples were collected are presented in Figure 5–2.

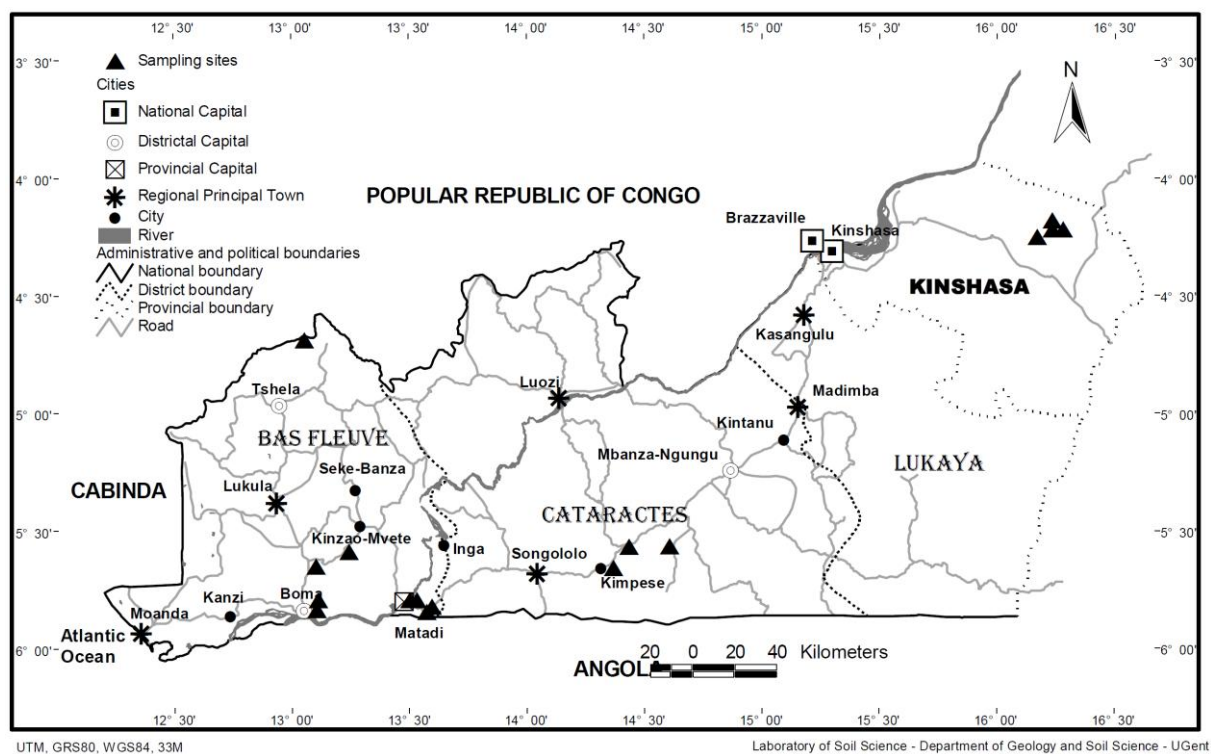


Figure 5–2. Location of the sampling sites in Lower Congo.

5.2.2 Evaluation dataset

The evaluation of the PTFs retained in this study was based on a dataset of soil samples collected during the soil survey of Lower Congo at the end of the 1980s and kept in the Laboratory of Soil Science at Ghent University (Belgium). The dataset constitutes of 48 disturbed samples and 12 undisturbed samples (mainly sandy soils) collected from the first four horizons of 15 representative soil profiles. Samples were taken at various depths up to 1.50 m following the subdivision in horizons of each profile. These profiles were mainly located in the regions of Bas-Fleuve, Cataractes and Kinshasa (the Bateke Plateau) (Figure 5–2). The profiles represent seven different Reference Soil Groups derived from five parent rocks (basic, calcareous, igneous and metamorphic rocks, micaschists and sandy materials). Soils were classified as Cambisols, Podzols, Phaeozems, Luvisols, Lixisols, Acrisols and Ferralsols (IUSS Working Group WRB, 2006) and a more detailed classification of these soils has been given in Chapter 2 (see Table 2–1). Ferralsols are dominant and represent 40% of the soil samples in the dataset. The samples analysed cover a wide range of textural classes as can be seen in Figure 5–3. Table 5–1 summarizes physical and chemical properties of the soil samples that are typically used in different PTFs.

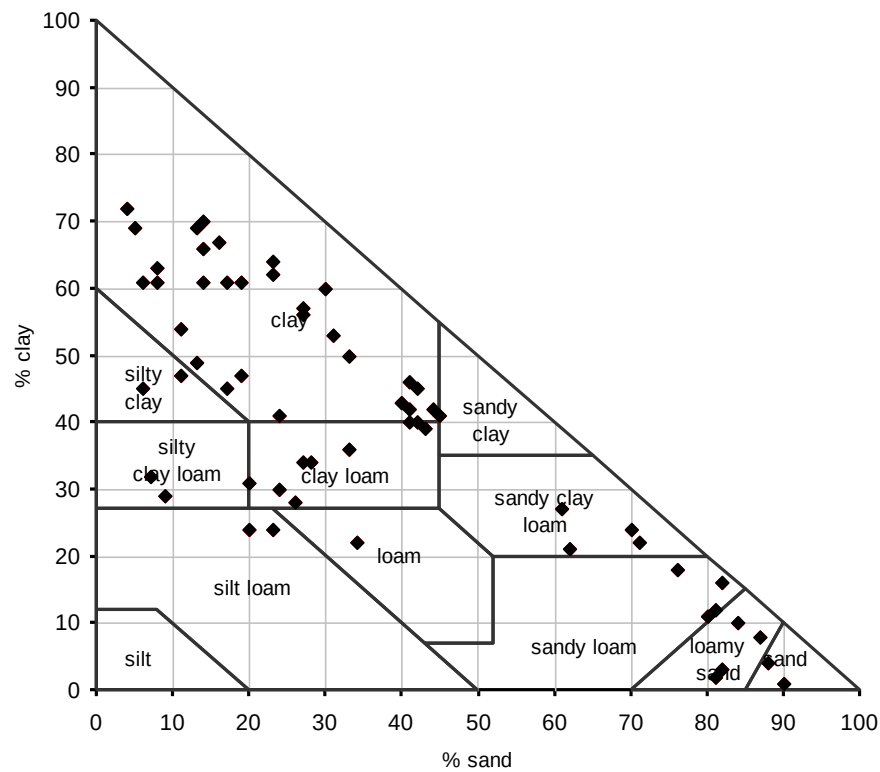


Figure 5–3. Variation of clay (0–2 µm), silt (2–50 µm) and sand (50–2000 µm) content of 60 selected soil samples of the Lower Congo dataset.

Table 5-1. Descriptive statistics of some physical and chemical soil properties of the Lower Congo soil dataset.†

Descriptive statistics	Clay	Silt	Sand	pH-H ₂ O	OC	CEC	DCB-Fe	DCB-Al	BD	W _{-33kPa}	W _{-1500kPa}	θ _{-33kPa}	θ _{-1500kPa}
Mean	40.0	24.0	40.0	5.6	1.18	10.23	6.15	1.17	1.17	0.233	0.157	0.262	0.178
Median	41.0	18.0	27.0	5.3	0.74	8.66	5.99	0.95	1.14	0.265	0.173	0.303	0.194
Std deviation	20.0	15.0	26.0	0.8	1.09	6.83	3.85	0.72	0.15	0.991	0.679	0.993	0.698
Minimum	1.0	2.0	4.0	4.4	0.09	0.68	0.02	0.03	0.85	0.015	0.011	0.026	0.017
Maximum	72.0	62.0	90.0	7.9	5.36	28.04	13.45	3.23	1.68	0.391	0.266	0.385	0.286

† Clay, Silt and Sand are clay (0–2 µm), silt (2–50 µm) and sand (50–2000 µm) fractions (%), OC is organic carbon (%), pH-H₂O (1:2, soil:water ratio) is acidity (–), CEC is cation exchange capacity (cmol_c kg⁻¹ soil), DCB-Fe is dithionite-citrate-bicarbonate extractable iron (%), DCB-Al is dithionite-citrate-bicarbonate extractable aluminium (%), BD is bulk density (Mg m⁻³), w is gravimetric water content at a specific matric potential (kg kg⁻¹), θ is volumetric water content at a specific matric potential (m³ m⁻³).

Number of samples = 60.

5.2.3 Pedotransfer functions considered in this study

For this evaluation study, two main categories of PTFs were used: ‘temperate’ PTFs and ‘tropical’ PTFs. The selected PTFs are listed in Table 5–2. They include point-based PTFs and parameter-based PTFs (respectively Type 2 and Type 3 PTFs in Wösten et al., 2001) for gravimetric water content w and volumetric water content θ . Most of the PTFs evaluated in this study were derived from soils located in the humid tropics. The exceptions were the PTFs of Oliveira et al. (2002) developed mainly for soils occurring under semi-arid conditions in Brazil. The well-documented ‘temperate’ point-based PTFs of Gupta and Larson (1979) and Rawls and Brakensiek (1982), and parameter-based PTFs of Vereecken et al. (1989) and Schaap et al. (2001) were introduced for comparison purposes.

Table 5-2. List of selected PTFs derived from ‘tropical’ and ‘temperate’ soils datasets.

Source	Outputs	Geographical domain
‘Tropical’ PTFs		
Aina and Periaswamy (1985)	$\theta_{-33 \text{ kPa}} = 0.6788 - 0.0055*Sa - 0.0013*BD$ $\theta_{-1500 \text{ kPa}} = 0.00213 + 0.0031*Cl$	Nigeria
Arruda et al. (1987)	$w_{-33 \text{ kPa}} = 0.29*(Cl + Si) + 9.93$ $w_{-1500 \text{ kPa}} = 0.27*(Cl + Si) + 1.07$ with Si (2–20µm)	South-East Brazil
Dijkerman (1988)	$w_{-33 \text{ kPa}} = 0.3697 - 0.0035*Sa$ $w_{-1500 \text{ kPa}} = 0.0074 + 0.0039*Cl$	Sierra Leone
Hodnett and Tomasella (2002)	$\ln \alpha = -0.02294 - 0.03526*Si + 0.024*OC - 7.6E^{-3}*CEC - 0.11331*pH$ $\ln n = 0.62986 - 0.00833*Cl - 0.00529*OC + 0.00593*pH + 7E^{-5}*Cl^2 - 1.4E^{-4}*Sa*Si$ $\theta_s = 0.81799 + 9.9E^{-4}*Cl - 0.3142*BD + 1.8E^{-4}*CEC + 0.00451*pH - 5E^{-6}*Sa*Cl$ $\theta_r = 0.22733 - 0.00164*Sa + 0.00235*CEC - 0.00831*pH + 1.8E^{-5}*Cl^2 + 2.6E^{-5}*Sa*Cl$	World tropical soils
Lal (1978)	$\theta_{-33 \text{ kPa}} = 0.065 + 0.004*Cl$ $\theta_{-1500 \text{ kPa}} = 0.006 + 0.003*Cl$	Nigeria
Oliveira et al. (2002)	$w_{-33 \text{ kPa}} = 0.00333*Si + 0.00387*Cl$ $w_{-1500 \text{ kPa}} = 3.8E^{-4}*Sa + 0.00153*Si + 0.00341*Cl + 0.030861*BD$	North-East Brazil
Pidgeon (1972)	$w_{FC} = 0.0738 + 0.0016*Si + 0.003*Cl + 0.03*OC$ $w_{-33 \text{ kPa}} = (100*w_{FC} - 3.77)/95$ $w_{-1500 \text{ kPa}} = -0.0419 + 0.0019*Si + 0.0039*Cl + 0.009*OC$	Uganda
van den Berg et al. (1997)	$\theta_{-1500 \text{ kPa}} = 0.00334*Cl*BD + 0.00104*Si*BD$	World Oxisols and related soils

θ = volumetric water content at a specific matric potential ($m^3 m^{-3}$), w = gravimetric water content at a specific matric potential ($kg kg^{-1}$), Cl = clay (%), Si = silt (%), Sa = sand (%), BD = bulk density ($Mg m^{-3}$), OC = organic carbon (%), OM = organic matter (%), CEC = cation exchange capacity ($cmol_c kg^{-1}$ soil); α , n , θ_s and θ_r are the parameters of the van Genuchten (1980) equation.

Table 5–2. List of selected PTFs derived from ‘tropical’ and ‘temperate’ soils datasets (continued).

Source	Outputs	Geographical domain
‘Temperate’ PTFs		
Gupta and Larson (1979)	$\theta_{-33 \text{ kPa}} = 0.003075 \cdot \text{Sa} + 0.005886 \cdot \text{Si} + 0.008039 \cdot \text{Cl} + 0.002208 \cdot \text{OM} - 0.01434 \cdot \text{BD}$ $\theta_{-1500 \text{ kPa}} = -0.000059 \cdot \text{Sa} + 0.001142 \cdot \text{Si} + 0.005766 \cdot \text{Cl} + 0.002228 \cdot \text{OM} + 0.002671 \cdot \text{BD}$	Eastern and Central US
Rawls and Brakensiek (1982)	$\theta_{-33 \text{ kPa}} = 0.2576 - 0.002 \cdot \text{Sa} + 0.0036 \cdot \text{Cl} + 0.0299 \cdot \text{OM}$ $\theta_{-1500 \text{ kPa}} = 0.026 + 0.005 \cdot \text{Cl} + 0.0158 \cdot \text{OM}$	US
Schaap et al. (2001)	van Genuchten (1980) retention parameters generated by the Rosetta software with Sa, Si, Cl, BD as predictors	US
Vereecken et al. (1989)	$\ln \alpha = -2.486 + 0.025 \cdot \text{Sa} - 0.351 \cdot \text{OC} - 2.617 \cdot \text{BD} - 0.023 \cdot \text{Cl}$ $\ln n = 0.053 - 0.009 \cdot \text{Sa} - 0.013 \cdot \text{Cl} + 0.00015 \cdot \text{Sa}^2$ $\theta_s = 0.81 - 0.283 \cdot \text{BD} + 0.001 \cdot \text{Cl}$ $\theta_r = 0.015 + 0.005 \cdot \text{Cl} + 0.014 \cdot \text{OC}$	Belgium

θ = volumetric water content at a specific matric potential ($\text{m}^3 \text{m}^{-3}$), w = gravimetric water content at a specific matric potential (kg kg^{-1}), Cl = clay (%), Si = silt (%), Sa = sand (%), BD = bulk density (Mg m^{-3}), OC = organic carbon (%), OM = organic matter (%), CEC = cation exchange capacity ($\text{cmol}_c \text{kg}^{-1}$ soil); α , n , θ_s and θ_r are the parameters of the van Genuchten (1980) equation.

5.2.4 Routine soil analyses

The routine analyses were carried out by Baert (1995) and Marcelino (1995). The physico-chemical characteristics of all soil samples (fine earth) were determined using standard methods described in Van Ranst et al. (1999). During these analyses, PSD (by the pipette method of Köhn, 1929), pH-H₂O, CEC, OC (by the method of Walkley and Black, 1934), base saturation (BS), dithionite-citrate-bicarbonate extractable iron (DCB-Fe) and aluminium (DCB-Al) (by the method of Mehra and Jackson, 1960) were determined among other things.

5.2.5 Soil water retention

Most soil samples used in this study were disturbed (2 mm-sieved). Water retention was thus determined on repacked samples similarly as, e.g., Pidgeon (1972), De Jong et al. (1983), Aina and Periaswamy (1985), and Gupta and Larson (1979) when they were developing their PTFs. Since we used repacked samples of 20.35 cm³-sized cylindrical cores, we only determined water retention at -33 and -1500 kPa, using pressure chambers (Soilmoisture Equipment, Santa Barbara CA) and following the procedures described in Cornelis et al. (2005). In this range, water content in the soil is less influenced by soil structure and more by soil texture (Juo and Franzluebbers, 2003). As indicated above, a matric potential value of -33 kPa is often considered as close to field capacity (FC) conditions of many soils, whereas a value of -1500 kPa is close to the permanent wilting point (PWP) of many crops (Cassel and Nielsen, 1986). Many authors have considered these two points in the development of point-based PTFs. Recently, Minasny and Hartemink (2011) reported that in most soil databases from tropical countries, water content at -10 or -33 kPa was measured on disturbed samples. During the soil survey of Lower Congo at the end of the 1980s, most soil samples collected for lab analyses were disturbed and BD and hydraulic properties data were missing. As indicated by Minasny and Hartemink (2011), this situation is closely related to the two primary objectives of soil survey campaigns in the tropics: soil mapping and classification for which undisturbed samples were not required. Bell and Van Keulen (1995) mentioned that caution should be taken when using FC data derived from disturbed samples. Therefore, we took some cautionary measures in that sense. Few undisturbed samples were recently collected from the same or very close locations as the soil profiles used in this study. We measured water content at FC on these undisturbed samples and we compared them with FC water content data on repacked soil samples. A similar approach was followed by Buccigrossi et al. (2010) in their PTFs evaluation study. They used disturbed samples from the Apulia region in Italy to determine water content at FC (-33 kPa) and PWP (-1500 kPa). Subsequently, they collected undisturbed samples and conducted an experiment which clearly showed that there was no significant difference between the water content of undisturbed and sieved soil at FC and PWP. In the present study, we made the same observations for Lower Congo soils. Furthermore, to test the success of our repacking procedure, we compared BD data from the repacked soils to the ones from undisturbed samples and from the neighbouring region Congo-Brazzaville reported by Schwartz and Namri (2002). No significant differences were found in relation to water retention and this indicates that the repacking procedure used was satisfactory for the purpose of this study.

5.2.6 Evaluation methods

A suitable evaluation of predictive methods such as PTFs requires the use of several validation indices simultaneously (Whitmore, 1991; Cornelis et al., 2001; Huang et al., 2003; Moreels et al., 2003). The prediction quality of the PTFs was determined by three complementary indices: the mean difference (MD), the standard deviation of the differences (SDD) and the root mean square difference (RMSD):

$$MD = \frac{1}{N} \sum_{i=1}^N (W_{pi} - W_{mi}) \quad [5-1]$$

$$SDD = \sqrt{\frac{1}{N-1} \sum_{i=1}^N ((W_{pi} - W_{mi}) - MD)^2} \quad [5-2]$$

$$RMSD = \sqrt{\frac{1}{N} \sum_{i=1}^N (W_{pi} - W_{mi})^2} \quad [5-3]$$

where W_{pi} is the predicted gravimetric or volumetric water content for soil sample i (kg kg^{-1} or $\text{m}^3 \text{m}^{-3}$), W_{mi} the measured water content for soil sample i (kg kg^{-1} or $\text{m}^3 \text{m}^{-3}$), and N the number of samples. For further information about the aforementioned validation indices, one can refer to De Vos et al. (2005).

The aforementioned indices were calculated for -33 kPa and -1500 kPa respectively. For further information about these statistical indices, one can refer to De Vos et al. (2005). In model performance assessment, the overall prediction error can be considered as the main criterion of comparison. It is mathematically expressed by the (R)MSD. For large N , the MSD equals the sum of MD^2 (bias) and SDD^2 (variance) (De Vos et al., 2005):

$$MSD = MD^2 + \frac{N-1}{N} SDD^2 \quad [5-4]$$

Baker and Ellison (2008b) refer to this equation as the bias-variance trade-off (BVTO). Indeed, attempts to decrease the bias are likely to result in higher variance and vice versa. Therefore, a compromise must be found between the conflicting requirements of small variance and small bias to produce an optimal fitting. SPSS 16.0.2 (SPSS, Inc., Chicago) was used for all statistical calculations necessary for the evaluation.

Van den Berg et al. (1997) considered FC at -10 kPa in the development of their PTFs. Therefore, the first PTF of these authors will not be considered in the evaluation of water content prediction at FC, which we took at -33 kPa. Silt in the PTF of Arruda et al. (1987) was determined as the fraction $2-20$ μm . Since information on the 20 μm measurement was available in the Lower Congo dataset, a quick recalculation of percents of sand and silt to match the boundary used by Arruda et al. (1987) was possible. Therefore, the PTF was applied using the converted PSD without the need of using one of the PSD models tested in Chapter 4.

5.3 Results and discussion

5.3.1 Exploratory study on potential predictors of water retention parameters

Table 5-3 shows the soil property ranges of the datasets used by the various authors to develop their ‘tropical’ and ‘temperate’ PTFs. This is based on available information about the range of the applied soil properties for each PTF provided by the authors. Some authors have developed PTFs based on data of a specific region only, and intended for a regional application (Gijssman et al., 2002). Cornelis et al. (2001) and Ungaro et al. (2005) warned that the extrapolation of PTFs beyond the statistical limits of the calibration datasets and the geographical locations of the soils from which they were derived should always be preceded by a careful evaluation of their applicability to specific soil datasets.

Table 5-3. Soil property ranges of the calibration datasets used to develop the evaluated ‘tropical’ and ‘temperate’ pedotransfer functions (PTFs).

Pedotransfer functions	Clay	Silt	Sand	OC/OM†	BD
	(%)				(Mg m ⁻³)
‘Tropical’ PTFs					
Aina and Periaswamy (1985)	8–43	6–20	42–86	unknown	0.85–1.68
Arruda et al. (1987)	unknown	unknown	unknown	unknown	unknown
Dijkerman (1988)	unknown	unknown	unknown	unknown	unknown
Hodnett and Tomasella (2002)	0–95.4 (36.7)††	0.4–77.4 (24.2)	0.5–99 (39.2)	0–30.8 (1.4)	0.28–1.88 (1.20)
Lal (1978)	13–56	6–20	20–78	0.16–2.84	NA
Oliveira et al. (2002)	unknown	unknown	unknown	unknown	unknown
Pidgeon (1972)	unknown	unknown	unknown	unknown	unknown
van den Berg et al. (1997)	10–95	unknown	unknown	0.1–5.2	0.8–1.6
‘Temperate’ PTFs					
Gupta and Larson (1979)	0–65	1–72	5–98	0–23	0.74–1.74
Rawls and Brakensiek (1982)	1–93 (18)	1–93 (26)	1–99 (56)	0–7 (1)	0.70–2.09 (1.42)
Schaap et al. (2001)	0–89	6–89	0–100	unknown	0.46–1.79
Vereecken et al. (1989)	0–55 (11)	0–81 (37)	6–98 (52)	0–13 (2)	1.04–1.83 (1.47)

† OC applies for the ‘tropical’ PTFs whilst OM applies for the ‘temperate’ PTFs;

†† Arithmetic means are given in parentheses.

Table 5–4 shows the correlation matrix between the properties of the soils considered in this study. Four of the most promising predictors were identified for gravimetric water content at FC and PWP, namely sand, clay, BD, DCB-Fe. They showed high to very high Pearson correlation coefficients with gravimetric water content. The same four properties were found to show high correlations with volumetric water content at FC and PWP. However, DCB-Fe seemed to yield a relatively better correlation with volumetric water content than BD. As could be expected, sand and BD were negatively correlated with water content whilst clay was positively correlated. Significant correlations ($p < 0.01$) were also found with silt, CEC and DCB-Al. In contrast, OC, pH and BS were not significantly correlated with water content at –33 and –1500 kPa.

Table 5-4. Pearson's correlation matrix between the various potential predictors and measured soil gravimetric and volumetric water retention points (N=60).

	Clay	Silt	Sand	OC	BD	pH	CEC	BS	DCB-Fe	DCB-Al	W_{-33kPa}	$W_{-1500kPa}$	θ_{-33kPa}	$\theta_{-1500kPa}$
Clay	1.00													
Silt	0.09	1.00												
Sand	-0.81	-0.66	1.00											
OC	-0.13	0.28	-0.06	1.00										
BD	-0.57	-0.34	0.63	-0.39	1.00									
pH	0.00	0.42	-0.25	0.17	0.12	1.00								
CEC	0.07	0.69	-0.45	0.60	-0.36	0.50	1.00							
BS	0.06	0.40	-0.28	0.30	-0.03	0.82	0.61	1.00						
DCB-Fe	0.77	0.34	-0.79	0.01	-0.56	0.25	0.22	0.27	1.00					
DCB-Al	0.57	0.02	-0.44	0.11	-0.48	-0.07	0.12	-0.10	0.71	1.00				
W_{-33kPa}	0.73**	0.63**	-0.92**	0.23	-0.77**	0.08	0.54**	0.24	0.69**	0.45**	1.00			
$W_{-1500kPa}$	0.91**	0.40**	-0.93**	0.10	-0.73**	0.11	0.35**	0.20	0.83**	0.60**	0.91	1.00		
θ_{-33kPa}	0.76**	0.63**	-0.94**	0.10	-0.65**	0.17	0.54**	0.30*	0.70**	0.42**	0.97	0.90	1.00	
$\theta_{-1500kPa}$	0.93**	0.38**	-0.93**	-0.01	-0.62**	0.17	0.32*	0.24	0.83**	0.57**	0.87	0.98	0.91	1.00

* shows significant correlation with water content at 0.05 significance level

** shows significant correlation with water content at 0.01 significance level

Figure 5–4 suggests that water content at –33 kPa and –1500 kPa can be satisfactorily estimated by sand, clay or DCB-Fe as the only predictor. The rise of DCB-Fe as a potential predictor of water content is worth to be mentioned in our study. Sand, silt, clay, CEC and BD are used as predictors in most of the PTFs considered in this evaluation study. In published water retention PTFs for ‘tropical’ soils, DCB-Fe and/or DCB-Al have almost never been considered as predictors for water retention. To our knowledge, only some parameter-based PTFs developed by van den Berg et al. (1997) used DCB-Fe and DCB-Al as previously mentioned in Chapter 3. Van den Berg et al. (1997) indicated that DCB-Fe plays an important role in the formation of the typical ‘coffee powder’ structure of many Ferralsols. This can explain the better performance of their parameter-based PTFs which used DCB-Fe and DCB-Al as predictors. In their study, these PTFs exhibited lower residual variances and higher r^2 values than alternative PTFs, i.e. those developed by excluding them as predictors.

The significant correlation between water content at FC and PWP on the one hand and DCB-Fe on the other hand (r equal to 0.69 and 0.83, respectively) can be partially explained by the stage of weathering of certain soils in our dataset. Indeed, the Lower Congo dataset is dominated by Ferralsols (40% of the soil samples), which are characterized by an advanced stage of weathering. Baert (1995) explained that in the ultimate stage, the breakdown of both unstable primary and 2:1 minerals becomes complete and, irrespective to the nature of the parent rock, the soils have reached ‘the mineralogical weathering residual system’ predominated mostly by Al- and Fe-(oxy) hydroxides and kaolinite (mineralogical maturity).

Based on the present dataset, DCB-Fe is found to be much more linearly correlated with water content at PWP than at FC (Figure 5–4). More than 69% of the variability in water content at PWP could be solely explained by DCB-Fe. This shows that DCB-Fe can be an as valuable predictor of water retention as clay or sand content. On the other hand, DCB-Al was a less successful predictor than DCB-Fe. Schaap et al. (2001) developed hierarchical ANN PTFs based on the availability of the predictors. Therefore, when developing hierarchical PTFs based on more comprehensive datasets, DCB-Fe could be considered as one of the potential predictors which can improve the prediction of water content at –33 and –1500 kPa for ‘tropical’ soils.

It is worth mentioning that a good positive correlation was also found between clay content and DCB-Fe ($r = 0.77$). This may explain why other tropical PTFs that do not include DCB-Fe but that have clay content in their equations might still show good prediction performance when applied for Lower Congo soils. Future studies should determine the contribution of DCB-Fe in the total variability of water content at FC and PWP when clay is also present as a predictor.

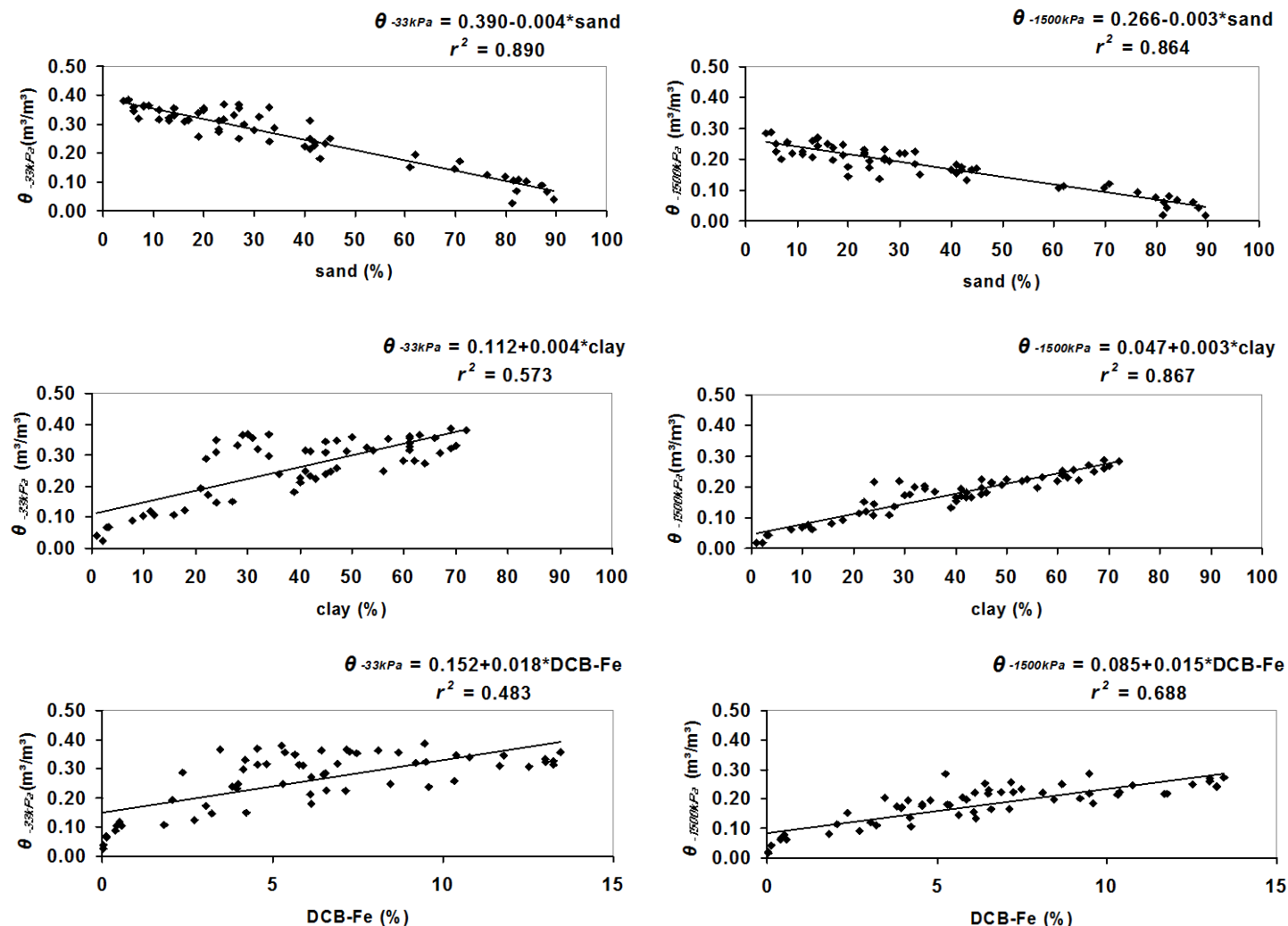


Figure 5–4. Regression equations and coefficients of determination (r^2) describing the relation between sand, clay, DCB-Fe contents (%) with soil water content ($m^3 m^{-3}$) of Lower Congo soils at –33 kPa and –1500 kPa.

5.3.2 Evaluation of the selected pedotransfer functions

The performance of the selected PTFs in predicting gravimetric and volumetric water content at FC and PWP of soils in Lower Congo is listed in Table 5–5 and Table 5–6. The PTFs considered in this study had contrasting prediction performance. The RMSD was used as the main evaluation criterion in this study. Indeed, it expresses the overall prediction error of a model. For gravimetric PTFs, the best PTFs at –33 kPa (FC) were Oliveira et al. (2002) with $\text{RMSD} = 0.0395 \text{ kg kg}^{-1}$ and Dijkerman (1988) with $\text{RMSD} = 0.0397 \text{ kg kg}^{-1}$. At –1500 kPa (PWP), the best PTFs were Arruda et al. (1987) with $\text{RMSD} = 0.0218 \text{ kg kg}^{-1}$ and Pidgeon (1972) with $\text{RMSD} = 0.0278 \text{ kg kg}^{-1}$ while the PTFs of Oliveira et al. (2002) performed here the worst with $\text{RMSD} = 0.0687 \text{ kg kg}^{-1}$. Therefore, the uncertainty in terms of RMSD can considerably vary from one gravimetric PTF to the other.

For volumetric PTFs, the best PTFs at FC were Hodnett and Tomasella (2002) followed closely by Schaap et al. (2001). The lowest performance was realized by Gupta and Larson (1979) and Aina and Periaswamy (1985). At PWP, the tropical PTF of van den Berg et al. (1997) performed the best. The temperate PTFs performed the worst and the PTFs of Schaap et al. (2001), Hodnett and Tomasella (2002) and Aina and Periaswamy (1985) showed average performance results.

Apart from the equations of Oliveira et al. (2002) and those of Hodnett and Tomasella (2002), the RMSD values of the different PTFs were lower in the prediction of gravimetric and volumetric water content at PWP than at FC. The Oliveira et al. (2002) equations presented a considerable increase of MD from $0.0021 \text{ m}^3 \text{ m}^{-3}$ at FC to $0.0654 \text{ m}^3 \text{ m}^{-3}$ at PWP. Likewise, the parameter-based PTF of Hodnett and Tomasella (2002) presented a slight underestimation of volumetric water content at FC ($\text{MD} = -0.0069 \text{ m}^3 \text{ m}^{-3}$) and a strong overestimation at PWP ($\text{MD} = 0.0397 \text{ m}^3 \text{ m}^{-3}$). However, in the evaluation study of Tomasella and Hodnett (2004), the general trend of lower RMSD values at PWP compared to FC conditions was observed for all the PTFs without any exception. Recently, Vereecken et al. (2010) also found that RMSD values at –1500 kPa are generally lower than at –33 kPa for ‘temperate’ soils.

Table 5-5. Validation indices of the gravimetric PTFs as computed on the complete evaluation dataset.†

PTF	<i>W–33 kPa</i>			<i>W–1500 kPa</i>		
	MD	SDD	RMSD	MD	SDD	RMSD
	kg kg^{-1}			kg kg^{-1}		
A	0.0212	0.0493	0.0533	-0.0028	0.0218	0.0218
D	0.0111	0.0385	0.0397	0.0055	0.0319	0.0321
L	-0.0374	0.0651	0.0746	-0.0519	0.0260	0.0580
O	0.0021	0.0398	0.0395	0.0654	0.0210	0.0687
P	0.0093	0.0443	0.0449	0.0127	0.0249	0.0278

† A: Arruda et al. (1987); D: Dijkerman (1988); L: Lal (1978); O: Oliveira et al. (2002); P: Pidgeon (1972).

Table 5-6. Validation indices of the volumetric PTFs as computed on the complete evaluation dataset.†

PTF	MD	$\theta_{-33 \text{ kPa}}$		MD	$\theta_{-1500 \text{ kPa}}$	
		SDD	RMSD		SDD	RMSD
		$\text{m}^3 \text{ m}^{-3}$			$\text{m}^3 \text{ m}^{-3}$	
AP	0.2176	0.0604	0.2257	-0.0518	0.0257	0.0577
GL	0.2990	0.0445	0.3023	0.0856	0.0538	0.1009
HT	-0.0069	0.0337	0.0341	0.0397	0.0216	0.0450
RB	0.1289	0.0679	0.1454	0.0804	0.0446	0.0918
SC	0.0373	0.0440	0.0574	-0.0286	0.0277	0.0397
VB	NA	NA	NA	0.0020	0.0216	0.0216
VE	0.1016	0.0580	0.1168	0.0857	0.0550	0.1016

† AP: Aina and Periaswamy (1985); GL: Gupta and Larson (1979); HT: Hodnett and Tomasella (2002); RB: Rawls and Brakensiek (1982); SC: Schaap et al. (2001); VB: van den Berg et al. (1997); VE: Vereecken et al. (1989). NA: Not applicable.

Graphs of MD^2 (bias error) versus SDD^2 (variance error) are shown in Figure 5–5 for the gravimetric PTFs and in Figure 5–6 for the volumetric PTFs. As mentioned earlier, the best models are the ones which show the lowest overall error, i.e. the best BVTO. They are located near the origin of the graph with the lowest (R)MSD values (De Vos et al., 2005). Regardless of the matric potential, the gravimetric PTFs of Pidgeon (1972) and Dijkerman (1988) have the lowest average overall error amongst all gravimetric PTFs (Figure 5–5). They present thus the best BVTO compared to other PTFs. This is particularly true for the PTFs of Dijkerman (1988) which gave good results in prediction of gravimetric water content at both FC and PWP.

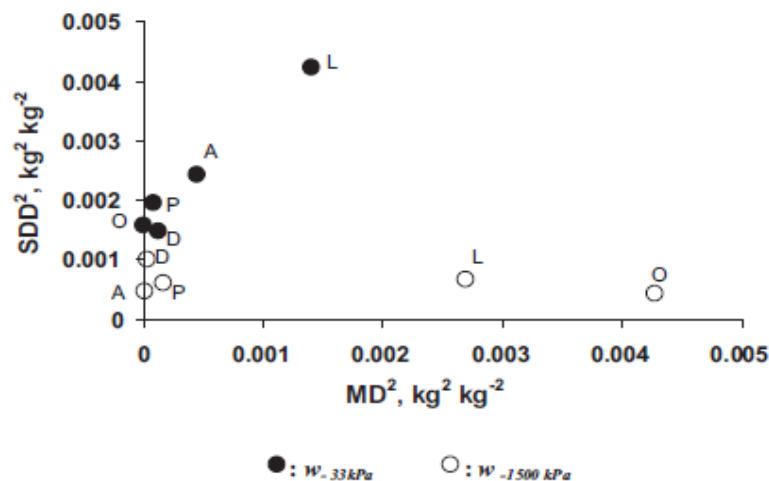


Figure 5–5. Mean difference squared (MD^2) versus standard deviation of the differences squared (SDD^2) of the gravimetric pedotransfer functions. A: Arruda et al. (1987); D: Dijkerman (1988); L: Lal (1978); O: Oliveira et al. (2002); P: Pidgeon (1972).

The best models to predict volumetric water content at FC and PWP in terms of BVTO are Hodnett and Tomasella (2002) and Schaap et al. (2001) as depicted in Figure 5–6. The van den Berg et al. (1997) PTF which predicts water content at PWP, but not at -33 kPa – note that they considered FC at -10 kPa – also performs very well. The PTFs showed more variations in their bias errors (MD^2) at -33 kPa . However at -1500 kPa , the bias errors yielded by these PTFs are more similar and of lower values. The precision errors (SDD^2) were quite variable at both matric potentials. Additionally, the

best PTFs in this study have also bias errors that are small enough to be considered as negligible in practice.

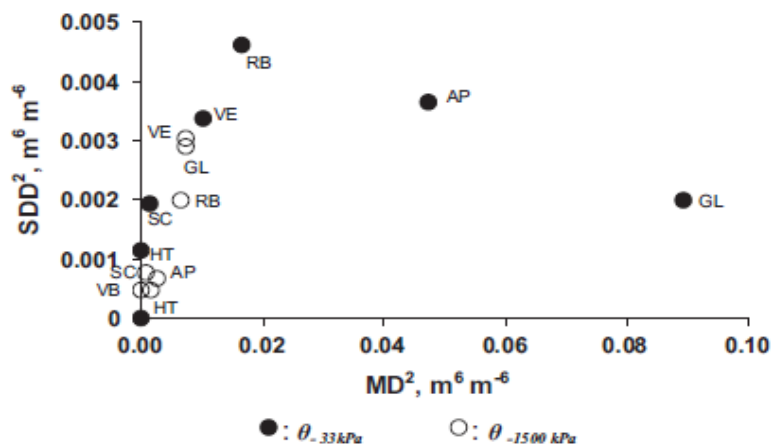


Figure 5–6. Mean difference squared (MD^2) versus standard deviation of the differences squared (SDD^2) of the volumetric pedotransfer functions. AP: Aina and Periaswamy (1985); GL: Gupta and Larson (1979); HT: Hodnett and Tomasella (2002); RB: Rawls and Brakensiek (1982); SC: Schaap et al. (2001); VB: van den Berg et al. (1997); VE: Vereecken et al. (1989).

Although most ‘temperate’ PTFs showed a poor performance compared to ‘tropical’ PTFs in estimating water content, it is remarkable to see that the ANN PTF by Schaap et al. (2001) showed a good performance for soils in Lower Congo. This is probably due to its flexibility and ability to mimic the complex behaviour of soils (Pachepsky and Schaap, 2004). The PTFs of Gupta and Larson (1979) and Rawls and Brakensiek (1982) did not perform better than any ‘tropical’ PTF. This is probably due to the differences in soil properties and in the mineralogy between the datasets used to develop these PTFs and the Lower Congo dataset. The PTF of Vereecken et al. (1989) still yielded a very high MD value which shows a notable overprediction of water content at PWP. This indicates that the range in clay content was not the major cause of its low predictive capability. Many authors, as Medina et al. (2002), indicated that clay type plays a vital role in the retention and transmission properties of a given soil. This is why soils in the humid tropics can have a much lower capacity to retain water than soils in the temperate regions with the same clay content but with a different type of clay.

Soils in Lower Congo are highly weathered soils principally dominated by kaolinite which is a non-expanding type clay mineral. In contrast, montmorillonite is a clay mineral mainly occurring in soils from temperate areas which can swell by absorbing substantial amounts of water in its interlamellar space (Van Ranst, 1997). This means that using these ‘temperate’ PTFs in models for agricultural, hydrological and ecological studies under the humid tropics can substantially reduce the quality of the modelling results. This can bias farmers’ decisions or experts’ recommendations in agricultural water management issues like irrigation scheduling, water management practices in rainfed farming or assessment of the risks of soil pollution in a watershed. The behaviours of all the PTFs in terms of measured versus predicted water contents for soils in Lower Congo are further illustrated in Figure 5–7 for gravimetric PTFs and in Figure 5–8 for volumetric PTFs.

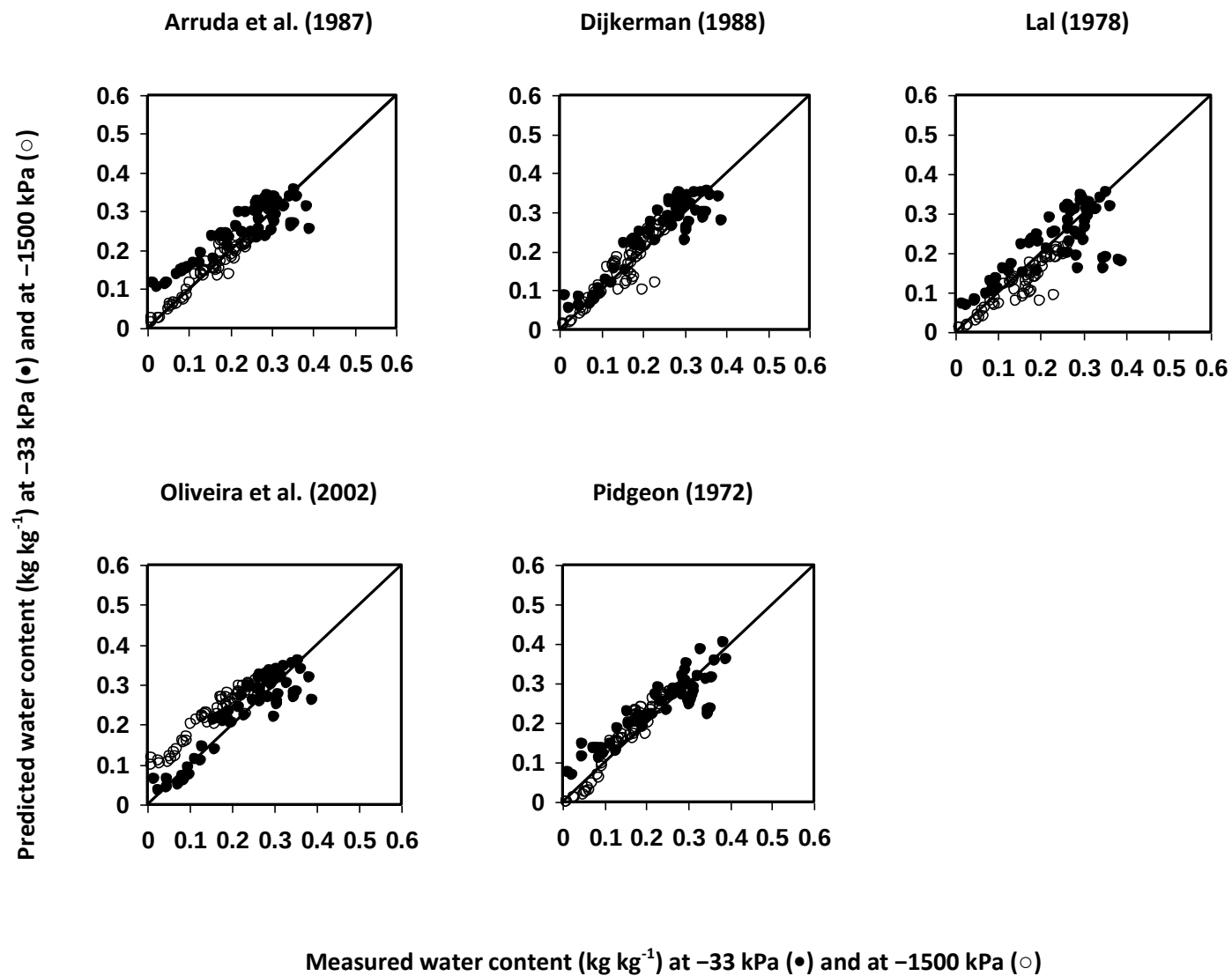
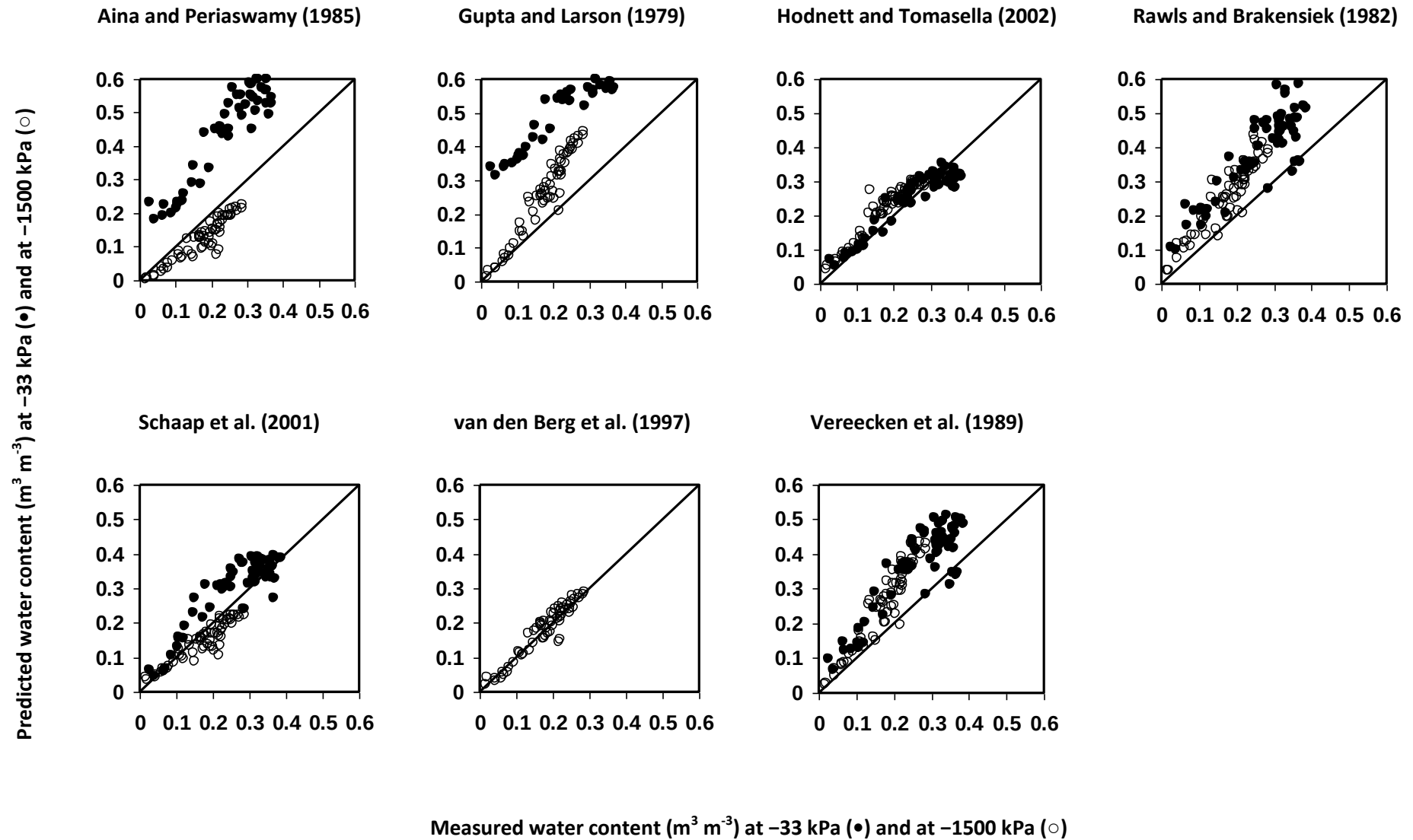


Figure 5–7. Measured versus predicted soil gravimetric water contents at -33 kPa and -1500 kPa

Figure 5–8. Measured versus predicted soil volumetric water contents at -33 kPa and -1500 kPa

5.4 Conclusions

The PTFs of Pidgeon (1972) and Dijkerman (1988) were the best amongst the gravimetric PTFs with better BVTO. They presented good results in prediction of gravimetric water content at both FC (RMSD = 0.0449 and 0.0397 kg kg⁻¹, respectively) and PWP (RMSD = 0.0278 and 0.0321 kg kg⁻¹, respectively), particularly the PTFs of Dijkerman (1988). Other gravimetric PTFs showed contrasting performance at -33 kPa and at -1500 kPa. The 'tropical' PTFs of Hodnett and Tomasella (2002) with RMSD = 0.0341 m³ m⁻³ at -33 kPa and 0.0450 m³ m⁻³ at -1500 kPa and van den Berg et al. (1997) with RMSD = 0.0216 m³ m⁻³ at -1500 kPa gave the best results in prediction of volumetric water content. The 'temperate' PTFs performed the worst for the Lower Congo soil dataset at -33 kPa and -1500 kPa. The notable exception was the 'temperate' PTFs of Schaap et al. (2001) which performed relatively well at both FC (RMSD = 0.0574 m³ m⁻³) and PWP (RMSD = 0.0397 m³ m⁻³). Estimates of water content by 'temperate' PTFs may induce errors in watershed models used in various agricultural studies under the humid tropics. Large discrepancies in the derived soil hydraulic data can substantially reduce the quality of the modelling results particularly in regions where soils may have been formed and evolved in similar climatological and pedological conditions as soils from Lower Congo.

However, further studies are recommended based on a larger dataset of undisturbed soils of Lower Congo. These studies can include the development of specific PTFs for water retention, K and BD as this information is generally lacking in soil datasets of the tropics. In the formulation and implementation of new PTFs, the potential of promising predictors such as DCB-Fe can be further assessed since data on DCB-Fe are generally available in various soil survey reports of Lower Congo. A more advanced step would be a functional evaluation of the newly developed PTFs in soil water management models such as SPAW or AquaCrop.

6 Prediction of water retention of soils from the humid tropics by the nonparametric k-nearest neighbour approach

This chapter is based on Botula, Y.-D., Nemes, A., Mafuka, P., Van Ranst, E., Cornelis, W.M., 2013. Prediction of water content of soils from the humid tropics by the non-parametric k-nearest neighbor approach (article published in Vadose Zone Journal doi:10.2136/vzj2012.0123).

6.1 Introduction

The unsaturated soil hydraulic functions are important parameters in many pedological, hydrological, ecological, and agricultural studies (Rajkai et al., 2004); however, direct measurements of such parameters are still expensive and time consuming, especially for studies at a regional scale (Vereecken, 1995; Pachepsky et al., 2006; Guber et al., 2006). Medina et al. (2002) stated that in developing countries, there are additional problems associated with this task, ranging from personnel training to acquisition of the necessary equipment. Therefore, an attractive alternative to the direct and often cumbersome measurements of soil hydraulic properties is their estimation by so-called pedotransfer functions. Bouma (1989) described the term pedotransfer function (PTF) as ‘translating data we have into what we need’. Pedotransfer functions thus relate more easily measurable soil data or other data routinely measured or registered in soil surveys with hydraulic parameters in a statistical sense (Bouma and van Lanen, 1987; Bouma, 1989; van den Berg et al., 1997).

Another alternative to obtain estimates or approximates of hydraulic properties is inverse modelling. Inverse procedures have the potential to yield information about soil hydraulic conductivity and water retention across a wide range of matric potentials from a single infiltration experiment (Schwartz and Evett, 2002). Briefly, the multistep outflow method applies an inverse modeling technique for indirect estimation of both water retention and hydraulic conductivity curves in a single transient drainage experiment (van Dam et al., 1994). The soil hydraulic parameters of an analytical function for the soil water retention curve (SWRC) (e.g., van Genuchten, 1980) or for hydraulic conductivity (K) (e.g., Mualem, 1976) are determined by matching experimental observations of transient water flow with numerical modelling results. In simple words, the estimated parameters are the solutions of an inverse problem. The latter results in determining causes that are unknown a priori based on observations of their effects. Hopmans et al. (2002) presented a comprehensive review of inverse modelling for estimation of soil hydraulic properties, including one-step and multistep methods. While this technique can yield a rather accurate set of effective soil hydraulic properties, its feasibility is limited for large-scale applications or for use in areas or countries with scarce resources.

When applying pedotransfer modelling or inverse modelling to obtain estimates or approximates of hydraulic properties, we should bear in mind that soils from tropical regions are vastly different from soils from temperate regions (e.g., van den Berg et al., 1997; Hodnett and Tomasella, 2002; Minasny and Hartemink, 2011). In Chapter 5, we evaluated the ability of some selected PTFs to predict the water content (θ) at -33 and -1500 kPa matric potentials for a limited dataset of soils from Lower Congo. We found that the temperate-climate PTFs of Gupta and Larson (1979) and Rawls and Brakensiek (1982) largely overestimated water retention of soils in Lower Congo. These PTFs were derived based on temperate-climate soils from across the US. On the other hand, we demonstrated that the tropical-climate PTFs of Hodnett and Tomasella (2002) performed well compared with the temperate-climate PTFs. Hodnett and Tomasella (2002) used a part of the IGBP-DIS soil database obtained from ISRIC-World Soil Information in Wageningen, the Netherlands, to derive PTFs for predicting the four parameters of the van Genuchten (1980) equation. They referred to this development dataset as the IGBP/T dataset, which contained exclusively soils from tropical climates. In Chapter 5, we attributed the poor predictive performance of the ‘temperate’ PTFs to the differences in soil properties and mineralogy between the test dataset and the dataset used to develop the PTFs. We recommended that more efforts should be done to develop specific PTFs to predict water retention of soils in the tropics. As previously mentioned, Schaap (2005) wrote that ‘with the exception of a few studies, hydraulic data and corresponding indirect methods about tropical soils are a virtual *terra incognita*’. This situation has not changed much by today. Minasny and Hartemink (2011) also noted

that limited efforts have been devoted to the prediction of properties of soils in the tropics where the need for accurate and up-to-date soil property information is even more urgent than elsewhere. They identified various soil properties used to predict the SWRC in the tropics, such as sand, silt, clay, bulk density (BD), organic carbon (OC) or organic matter (OM), pH, cation exchange capacity (CEC), dithionite-citrate-bicarbonate extractable iron (DCB-Fe) and aluminium (DCB-Al), but finally selected soil texture, BD, and OC to develop PTFs to predict water content at -10 , -33 , and -1500 kPa. The development dataset and the validation dataset exclusively contained soils from the tropics. These soil datasets are also part of the international IGBP-DIS soil database obtained from ISRIC.

Despite the limited efforts in data collection and harmonization for soils from the humid tropics (where most of the developing countries are located) compared with temperate areas, large tropical soil databases will steadily grow. With the emergence of such large databases, classical statistical methods such as multiple linear regression (MLR) may have limitations because important trends may not be detected, whereas others may falsely be given much emphasis. Therefore, there is a need to promote data-mining or pattern-recognition techniques, which are flexible enough to handle huge amounts of data and detect important trends that may be hidden from classical statistical methods such as MLR.

Even though classic PTFs based on the MLR approach have been widely used to predict water retention in the past, PTFs based on pattern-recognition approaches have gained popularity. This is particularly because they have the advantage of including new soil information without the constraint of developing new equations to fit the new soil dataset. This flexibility in incorporating new soil data is highly beneficial in tropical regions, particularly for developing countries, where continuously developing soil databases are in high demand for pedological, agricultural, and ecological studies. Pattern-recognition techniques belong to the group of data-driven, data-mining, or machine-learning techniques, in contrast with MLR, which is based on predefined mathematical functions. Recently, three pattern-recognition techniques were used with success in studies related to unsaturated soil hydrology: artificial neural networks (ANN), support vector machines (SVM), and the k-nearest neighbour (k-NN) technique. Mucherino et al. (2009) provided an elaborated review of these data-mining techniques and on their application in various agriculture- and environment-related fields. For further information on the ANN and SVM techniques, see Hecht-Nielsen (1990), Haykin (1994), Vapnik (1995, 1998), and Noble (2006).

In this study, we used the k-NN technique, which has been considered one of the most attractive pattern-recognition algorithms by several researchers (e.g., Buishand and Brandsma, 2001; Bannayan and Hoogenboom, 2009). It has been referred to as a ‘lazy learning algorithm’ because it passively stores the data until the time of application. All calculations are performed in ‘real time’, i.e., only when estimations need to be generated. Application of the k-NN technique involves identifying and retrieving the instances most similar to the target object from the multidimensional feature (input variable) space of the set of stored instances and classifying the target object based on similarities in their input attributes using a predefined weighting scheme. More theoretical details on this similarity-based approach were given in Dasarathy (1991).

Nemes et al. (1999) used a k-NN variant—which they termed the similarity technique to estimate missing soil particle size distribution (PSD) points from other existing PSD points to harmonize data of the European HYPRES database (Wösten et al., 1999). Jagtap et al. (2004) used a k-NN technique to estimate the drained upper limit and lower limit of plant-available water from soil water retention data measured in situ. Nemes et al. (2006a) provided several examples of applications of the k-NN techniques in hydrologic simulation and developed another variant of the k-NN technique to estimate

soil water retention at two matric potentials. They also performed a detailed sensitivity analysis of this technique (Nemes et al., 2006b). The newly developed k-NN algorithm proved its robustness in different scenarios. Based on the satisfactory results of their k-NN algorithm, Nemes et al. (2008) developed a user-friendly software called k-Nearest to estimate $\theta_{-33\text{kPa}}$ and $\theta_{-1500\text{kPa}}$ with the option of estimating the uncertainty of the prediction using data resampling. Elshorbagy et al. (2010a,b) conducted a detailed study of the predictive capabilities of data-driven modelling techniques in hydrology and identified the k-NN technique as an attractive modelling technique for hydrologic applications because of its high level of flexibility due to reasons mentioned above. Nemes et al. (2006a) specifically referred to the k-NN method of working with patterns of similarities instead of fitting equations to data and its real-time application giving users the flexibility to alter the underlying data or the calculation scheme. Gharahi Ghehi et al. (2012) recently applied the k-NN approach to predicting the BD of Rwandese soils in the humid tropics.

When predicting hydraulic properties on the basis of existing databases for training data-driven models, Perkins and Nimmo (2009) stressed the necessity of high-quality databases. They indicated that an obvious problem occurs when the available database has few or no data for samples that are closely related to the region of interest. This is classically the case when a dataset of soils from temperate areas is used as a training dataset to predict the hydraulic properties of soils from tropical regions. In their sensitivity analysis, Nemes et al. (2006b) used separate datasets from the US, Europe, and Brazil and found that when using a dataset of ‘temperate’ soils as a training dataset to predict the water retention of ‘tropical’ soils from Brazil, estimations were significantly worse than for other examined dataset pairs, with bias errors amounting to an undesirable $0.10 \text{ m}^3 \text{ m}^{-3}$. As a point of future research, Nemes et al. (2006a) recommended testing the ability of the k-NN approach to predict soil water retention based on datasets from different regions of the world, but an application that uses an international collection of soils from the humid tropics is still lacking.

Point-estimation PTFs are usually limited to estimating only a few points on the water retention curve, most frequently two or three points. Among such applications are estimations using k-NN. In their application, Nemes et al. (2006a) predicted water content at -33 and -1500 kPa matric potentials using their k-NN variant and with a small number of input attributes: texture (sand + silt + clay, designated here as SSC), OM, and BD. Recently, Patil et al. (2012) used the k-NN software developed by Nemes et al. (2008) to estimate the $\theta_{-33\text{kPa}}$ and $\theta_{-1500\text{kPa}}$ of 157 swelling–shrinking soils in India to derive their available water capacity. These matric potentials were also used by numerous other studies (e.g., Givi et al., 2004; Reichert et al., 2009; Minasny and Hartemink, 2011). The rationale is that these two points are meant to be used as approximates to water retention at field capacity FC ($\theta_{-33\text{kPa}}$) and the permanent wilting point PWP ($\theta_{-1500\text{kPa}}$) to calculate the available water holding capacity or to parameterize bucket-type agronomic or water balance models. This raises two considerations that were of significance when initiating this study.

First, it is still debated what, if any, matric potential is a good representation of conditions at or near FC. It appears to be generally affected by a number of factors, among them the soil texture. Apart from field experiments (Ottoni Filho and Ottoni, 2010) and data-mining studies (Nemes et al., 2011), Twarakavi et al. (2009) also demonstrated this dilemma using inverse modelling. For ‘tropical’ soils, however, several researchers (e.g., Sharma and Uehara, 1968; Pidgeon, 1972; Babalola, 1979; Lal, 1978; Reichardt, 1988) have suggested that the water content at -10 kPa represents FC better than the water content at -33 kPa, which is more frequently adopted by researchers working with temperate-climate soils. The soil-water relation of well-aggregated kaolinitic soils under a tropical climate can be markedly different from that of soils with permanent-charge minerals in temperate regions. Heavy-

textured soils dominated by kaolinite and sesquioxides have SWRCs that in some respects resemble those of sandy soils (Sharma and Uehara, 1968), although they show higher porosity. In aggregated highly weathered soils (e.g., Ferralsols), water can reside in large interaggregate pores and fine intraaggregate pores. Under gravitational forces, water in the large pores moves rapidly and FC is attained at high matric potentials, generally between -10 and -15 kPa. Field capacity is attained at this high matric potential because the hydraulic conductivity at this potential is very low, much like that of a sandy soil. It may therefore be advisable to have information on the water content at higher matric potentials than -33 kPa when it comes to supporting studies in the humid tropics that concern the unsaturated zone. At the same time, according to the studies cited above, the water content at -1500 kPa can still be considered as an approximation of the PWP.

The second consideration is that when two or three points are estimated on the SWRC, it allows no or only limited (constrained) use of popular water retention models like the models of van Genuchten (1980) or Brooks and Corey (1964). Pedotransfer functions that estimate the parameters of such models offer a solution to this dilemma; however, it was found by Tomasella et al. (2003) that estimating SWRC points followed by curve fitting yielded more accurate results than estimating curve parameters and reading water content values at particular matric potentials off the fitted curve. Hence, we have chosen to estimate a number of water retention points that will facilitate the subsequent use of both point and parameterized SWRCs. Overall, to be able to fit a complete SWRC, six to eight measured or estimated water retention points are recommended because the SWRC models more commonly used (e.g., Brooks and Corey, 1964; van Genuchten, 1980) have four or more fitting parameters (Tomasella et al., 2000; Cornelis et al., 2005). Until now, no study has been published that estimates water content at more than two matric potentials using the k-NN method. Facilitated by the databases available for this study, we estimated up to eight SWRC points.

Therefore, the objectives of this study were (1) to apply a nonparametric approach to obtain estimations of water content of soils from a tropical region based on an international database of soils from the humid tropics and using an adaptation of the k-NN algorithm developed by Nemes et al. (2006a), (2) to test the ability of the k-NN algorithm to predict several points on the SWRC (i.e., water content at eight different matric potentials) from the wet to the dry range simultaneously, (3) to use a range of input attributes and determine the influence of several combinations of input attributes on the ability of the k-NN approach to predict the water content at those matric potentials, (4) to investigate the importance of particular input attributes in the estimation of soil water content at low, intermediate, and high matric potentials, and (5) to compare the prediction performance of the proposed k-NN variant and two aforementioned MLR PTFs that were developed using datasets from the tropics, similarly extracted from the international IGBP database.

6.2 Materials and methods

6.2.1 Soil datasets

In this study, a dataset of 534 soils from tropical regions was used as the reference or training dataset for the k-NN estimations. These soil samples are part of the IGBP-DIS international database from ISRIC (Tempel et al., 1996). By tropical regions, we mean regions situated between 25° N and 25° S and mainly under a (sub)humid climate. Soils within the tropics but in temperate climates due to altitude or in dry areas were not included in the selected dataset. This ‘tropical’ dataset is referred to here as the IGBPTrop dataset. It contains highly weathered soils such as Ferralsols (20.4%), Acrisols (11.6%), and Nitisols (4.7%) and other soils like Cambisols (14.2%), Andosols (6.4%), Luvisols (6%), Gleysols (4.7%), Phaeozems (4.3%), Fluvisols (2.8%), Vertisols (2.6%), and Arenosols (2.4%) among

others (IUSS Working Group WRB, 2006). Undisturbed and disturbed soil samples were collected under different land uses and from various depths. The associated digital database contains, among other attributes, water content data at eight different matric potentials (0, -1, -3, -10, -20, -50, -250, and -1500 kPa). Tempel et al. (1996) provided the necessary references concerning the different analytical methods used to derive the soil physical and chemical properties recorded in the database.

A dataset of 139 soils from Lower Congo was used as an independent dataset to test the predictive ability of the k-NN approach. These soils were mainly highly weathered soils under the humid tropics classified as Ferralsols, Acrisols, and Nitisols (IUSS Working Group WRB, 2006), but other Reference Soil Groups such as Arenosols were also represented. The 139 selected soil samples were not part of the IGBP-DIS database. Undisturbed soil samples were collected in 100-cm³ Kopecky rings under different land uses (savannah, forest, agricultural fields, and old quarries) and at various depths in the soil profile. For the undisturbed samples, the SWRC data pairs were determined from the wet to the dry range at eight different matric potentials: -1, -3, -6, -10, -20, -33, -100, and -1500 kPa. The hanging water-column method was used for matric potentials between -1 and -10 kPa using the sand box apparatus (Eijkelkamp Agrisearch Equipment), whereas for matric potentials between -20 and -1500 kPa, pressure chambers (Soilmoisture Equipment) were used, following the procedures described by Cornelis et al. (2005). The coupled matric potential–water content pairs represent single measurements on single samples. Matric potentials at 0, -50, and -250 kPa used in the IGBP-Trop database were missing in the Lower Congo database. Therefore, they were derived by curve fitting as follows: (i) a continuous curve was fitted through the discrete set of measured (available) water retention points using the van Genuchten (1980) function, and (ii) fitted values of water contents at the missing matric potentials (0, -50, and -250 kPa) were calculated from the resulting continuous equation. The physicochemical characteristics of all soil samples (fine earth) were determined using standard methods described in detail by Van Ranst et al. (1999). During these analyses, the PSD (by the pipette method of Köhn, 1929), OC, pH, and CEC were determined on the same soil samples that were previously used for the SWRC measurements.

Soil properties selected for use in this study were the following: sand (50–2000 µm), silt (2–50 µm), and clay (< 2 µm) contents according to the USDA classification system (Soil Survey Staff, 1951), BD, OC, pH, CEC, and retained (volumetric) water content (θ) at eight different matric potentials: 0, -1, -3, -10, -20, -50, -250, and -1500 kPa. Any entries that showed obvious inconsistencies in physical or hydraulic data (e.g., sand + silt + clay $\neq$ 1; $[(1 - BD/2.65) - \theta_{0kPa}] < 0$; $\theta_{xkPa} < \theta_{ykPa}$ when $x > y$ kPa) were excluded from the reference or training dataset and the test dataset. Figure 6–1 shows the textural distribution of the IGBP-Trop and the Lower Congo datasets.

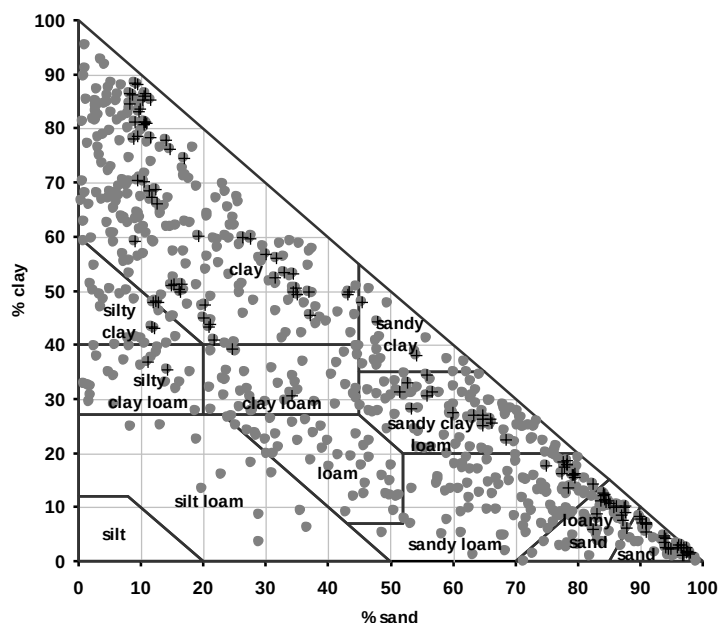


Figure 6–1. Variation of clay (0–2 μm), silt (2–50 μm) and sand (50–2000 μm) of 534 soil samples in the IGBP-Trop dataset (circles) and of 139 soil samples in the Lower Congo soil dataset (crosses).

6.2.2 The k-nearest neighbour technique

The k-NN algorithm used in this study was adapted from the variant developed by Nemes et al. (2006a). The same algorithm was used in this study but was expanded to use more input and output attributes, and the design parameters of the algorithm were re-evaluated for the current application. The implementation was done in the MATLAB R2010a environment (The MathWorks, Inc., Hill Drive Natick, MA).

6.2.2.1 Rationale

The k-NN technique does not use any predefined mathematical function to estimate a certain response attribute in the way classic MLR PTFs do. It does not appear to rely on any stringent assumptions about the underlying data and can adapt to any situation (Hastie et al., 2009). The k-NN approach consists of finding the k number of nearest neighbours from a reference dataset for each soil in the test dataset in terms of their selected input attributes. The similarity distance to the target soil is measured in terms of Euclidean distance after normalization and rescaling of the soil attribute data in the reference dataset following a specific procedure. This is done to ensure that different input attributes will receive equal weight. Soils in the reference dataset are sorted in ascending order of their (normalized) similarity distance to the target soil.

The number of selected nearest soil instances (k) also needs to be optimized following a specific procedure. Once the nearest neighbours are identified and sorted, distance-dependent weights are assigned to them and the response attribute is formulated and outputted as the weighted average of the response attributes of the selected nearest neighbours. More methodological and calculation details on the whole procedure are given below.

6.2.2.2 Selection of nearest neighbours to target soil

An external training (reference) dataset containing information on a wide variety of soils is searched for soils (instances) that are most similar to the target soil based on the selected input attributes or

features. Similarity between the target soils and the known instances is measured in terms of a metric considered here as the Euclidean distance:

$$d_i = \sqrt{\sum_{j=1}^x \Delta a_{ij}^2} \quad [6-1]$$

where d_i is the ‘distance’ of the i^{th} soil from the target soil and Δa_{ij} is the difference of the i^{th} soil from the target soil in the j^{th} soil attribute, x is the total number of soil attributes.

Soils of the reference dataset are sorted in ascending order of their distance to the target soil.

6.2.2.3 Normalization of soil data

Soils present some properties (attributes) that differ in their order of magnitude or range. For instance, a nonorganic soil can have 100% sand but should not have more than 18% OC (Soil Survey Staff, 1975). Therefore, a unit difference in OC is expected to be more significant than the same unit difference in sand content. Therefore, a normalization procedure was applied to the soil property data before they were used to calculate the Euclidean distance in Equation 6–1. Normalizing the soil attributes has the benefit of lowering bias toward one soil attribute or another. All input attributes were first transformed to temporary variables $a_{ij(temp)}$ with a distribution having zero mean and standard deviation of 1 by the following classic formula:

$$a_{ij(temp)} = \frac{(a_{ij} - \bar{a}_j)}{\sigma(a_j)} \quad [6-2]$$

where a_{ij} is the value of the j^{th} attribute of the i^{th} soil, and $\bar{a}_j$ and $\sigma(a_j)$ are the mean and standard deviation of the observed values of the j^{th} attribute in the reference dataset.

Second, the difference between the minimum and maximum of these temporary variables was then examined to identify the soil attribute that showed the widest range of transformed (temporary) values. This allowed a scaling of the temporary variables to obtain zero mean and the same minimum–maximum range in the data of all attributes:

$$a_{ij(trans)} = \frac{a_{ij(temp)} (\text{Max}(\text{range}(a_{j=1(temp)}), \dots, \text{range}(a_{j=x(temp)})))}{\text{range}(a_{j(temp)})} \quad [6-3]$$

where $a_{j(temp)}$ is the data of the j^{th} soil attribute normalized using Equation 6–2 and $a_{ij(trans)}$ is the final transformed value of the j^{th} attribute of the i^{th} soil. Eventually, $a_{ij(trans)}$ values derived from Equation 6–3 were used as input in our k-NN algorithm.

6.2.2.4 Application of a distance-dependent weighting system

A weighting procedure that accounts for the distribution of the distances of the selected k neighbours from the target soil was applied. Weights of each selected neighbour were computed as

$$w_i = \frac{d_{i(rel)}}{\sum_{i=1}^k d_{i(rel)}} \quad [6-4]$$

where k is the number of neighbours selected, w_i is the weight associated to the i^{th} nearest neighbour, and $d_{i(rel)}$ is the relative distance of the i^{th} selected neighbour calculated as

$$d_{i(rel)} = \left(\frac{\sum_{i=1}^k d_i}{d_i} \right)^p \quad [6-5]$$

where d_i is the distance of the i^{th} selected neighbour computed using Equation 6-1 and p is a power term to account for different possible weight–distance relationships.

Therefore, the predicted water retention at a given matric potential corresponds to the (distance-dependent) weighted sum of observed water retention values of the selected nearest neighbours.

6.2.3 Design parameters k and p for the k-nearest neighbour algorithm

There are two design parameters of the k-NN algorithm that were used, namely the k and p terms. The k term refers to the number of similar soils to be selected from the reference dataset to estimate the output attributes for each target soil, while the p term determines the weight–distance relationship that determines the contribution of each of the k reference samples to the estimation of the output attribute, depending on their degree of similarity to the target soil.

Nemes et al. (2006a) indicated that the best combination of k and p values, i.e., the one leading to the lowest overall prediction error (expressed by the root mean square difference [RMSD]) should be selected and that such a choice may depend on the size of the reference dataset. They tested this assumption on different dataset sizes, i.e., $N_r = 100, 200, 400$, and 800 , and derived two different functions for k and p that are dependent on the size N_r of the reference dataset:

$$k = 0.655 N_r^{0.493} \quad [6-6]$$

$$p = 0.767 N_r^{0.049} \quad [6-7]$$

They warned, however, that the relationships among N_r , k , and p in Equations 6-6 and 6-7 were set empirically and may not be optimal for other datasets. They recommended testing the settings of the k and p parameters for particular applications.

In this study, we reoptimized the two parameters using an approach similar to the one used by Nemes et al. (2006a). We determined what influence, if any, different k and p values had on the prediction performance of the k-NN algorithm in a tropical context, i.e., when soils from Lower Congo were used as a test dataset and the international IGBP-Trop dataset as a training dataset. To avoid possible bias toward one or another set of inputs, all predetermined input variables (i.e., SSC + BD + OC + pH + CEC) to estimate all eight water retention points as outputs were considered. Then, all the corresponding RMSDs were computed and plotted for a visual examination, and the best combination of k and p values was selected for this particular application. Because the difference in RMSDs between two subsequent p values was rather small, we decided to consider a change in p from 0.5 to 2.5, in increments of 0.5, whereas the values of k were changed from 0 to 50 in increments of 1. The optimized combination of k and p was then used in further calculations.

6.2.4 Ensemble of k-nearest neighbour estimations

We experimented with the influence of the reference dataset size, similarly to Nemes et al. (2006a), and so samples were drawn to be included in development or reference datasets of 100, 200, 300, 400, and 534 samples (i.e., all samples with available data). All random data selections were repeated 100 times to allow the development of an ensemble of water retention estimations. For each dataset size, the development or reference dataset was randomly sampled 100 times at an 80% resampling rate, i.e., a different subsample representing 80% of the reference dataset was used in each of the 100 replicates.

An ensemble of estimations has numerous advantages: the impact of any single replicate (i.e., any particular dataset division) on the final estimation results can be minimized when a sufficiently large number of replicates are used. Moreover, generation of an ensemble of estimations allows the quantification of the uncertainty of estimates, which can be used in statistical analyses or be used as input in simulation models. Quantification of uncertainty in estimates of soil hydraulic properties by PTFs and its effects in various simulation models has been studied by several researchers (Finke et al., 1996; Nemes et al., 2003; Deng et al., 2009; Loosvelt et al., 2011; Moeys et al., 2012), who indicated that the uncertainty associated with hydraulic PTFs should be taken into account when evaluating simulation results yielded by a given model.

In this study, we found empirically that 100 replicates were sufficient to make the effect of any single replicate on the estimations negligible. Therefore, in this study we used 100 replicates in the algorithm and any statistical measures were computed based on those 100 replicates; however, we also examined the minimum (optimized) number of replicates for each of the different dataset sizes ($N_r = 100, 200, 300, 400$, and 534).

6.2.5 Input and output attributes used

In this study, we selected a wide range of soil attributes as potential predictors. Not only have these soil properties been used by several researchers for the determination of ‘tropical’ PTFs but they are also important to characterize soils in the (sub)humid tropics: sand, silt, clay, BD, OC, pH, and CEC. Fourteen different combinations of these input attributes were considered to generate estimations in a hierarchical structure in order to evaluate which, if any, of the variable combinations would yield systematically better estimates. The output attributes were water content at eight different matric potentials, namely at 0, -1, -3, -10, -20, -50, -250, and -1500 kPa. This means that we estimated more water retention points, simultaneously, in the wet, the intermediate, and the dry range of the SWRC.

6.2.6 Evaluation criteria

Three statistical measures were selected to assess the predictive ability of the k-NN algorithm at a given matric potential: the mean difference (MD), the root mean square difference (RMSD), and the coefficient of determination (R^2):

$$MD = \frac{1}{N_t} \sum_{i=1}^{N_t} (\theta_{p_i} - \theta_{m_i}) \quad [6-8]$$

$$RMSD = \sqrt{\frac{1}{N_t} \sum_{i=1}^{N_t} (\theta_{p_i} - \theta_{m_i})^2} \quad [6-9]$$

$$R^2 = \frac{\left(\sum_{i=1}^{N_t} (\theta_{p_i} - \overline{\theta_{p_i}})(\theta_{m_i} - \overline{\theta_{m_i}}) \right)^2}{\sum_{i=1}^{N_t} (\theta_{p_i} - \overline{\theta_{p_i}})^2 \sum_{i=1}^{N_t} (\theta_{m_i} - \overline{\theta_{m_i}})^2} \quad [6-10]$$

where θ_{p_i} is the predicted volumetric water content for the i^{th} soil sample ($\text{m}^3 \text{m}^{-3}$), θ_{m_i} is the measured volumetric water content for the i^{th} soil sample ($\text{m}^3 \text{m}^{-3}$), and N_t is the number of samples in the test dataset.

6.2.7 Comparison with two published ‘tropical’ pedotransfer functions

The prediction performance of the proposed k-NN approach was compared with the prediction performance of the MLR PTFs of Hodnett and Tomasella (2002) and Minasny and Hartemink (2011) based on their RMSD values. As mentioned above, the PTFs of Hodnett and Tomasella (2002) predict the parameters of the van Genuchten (1980) equation based on basic soil properties (texture, BD, OC, pH, and CEC). Therefore, they allow the calculation of water content at any given matric potential. On the other hand, the PTFs of Minasny and Hartemink (2011) predict water content from texture, BD, and OC at three matric potentials: -10 , -33 , and -1500 kPa. In the present study, only the results for -10 and -1500 kPa were considered in the comparison with the k-NN approach because water content at -33 kPa is lacking in the IGBP-Trop dataset.

6.3 Results and discussion

Box plots of the selected soil attributes for the reference or training dataset (IGBP-Trop) and for the test dataset (Lower Congo) are given in Figure 6–2. Based on these soil attributes, it can be seen that both the reference and test datasets contained data of a wide range of soils.

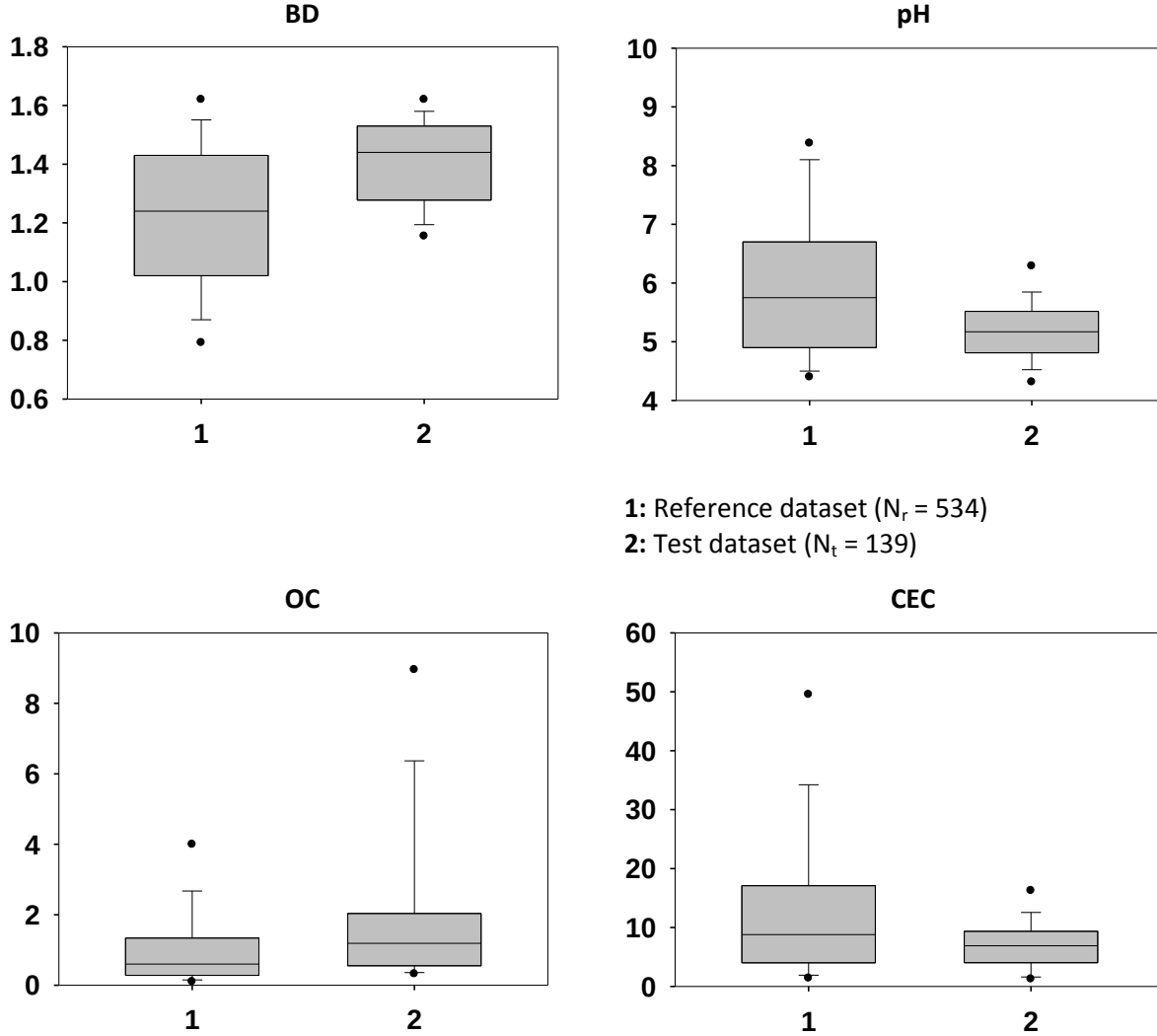


Figure 6–2. Box plots of some physical and chemical properties of the soils of (1) IGBP-Trop (reference dataset) and (2) Lower Congo (test dataset). BD is bulk density (Mg m^{-3}), OC is organic carbon content (%) and CEC is cation exchange capacity ($\text{cmol}_c \text{ kg}^{-1}$ soil).

6.3.1 Ensembles of k-nearest neighbour estimations

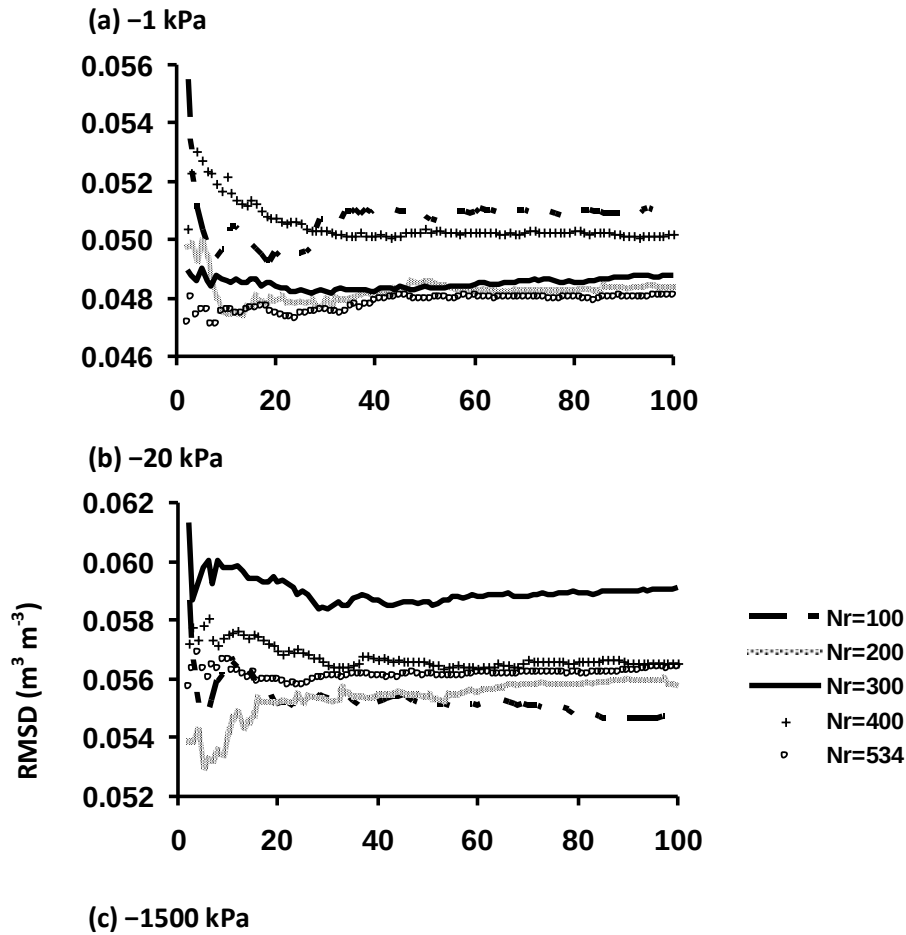
To find a minimum number of ensembles to obtain a stable RMSD based on the IGBP-Trop dataset, we plotted the running (cumulative) RMSD values against the total number of ensemble members after each replicate dataset had been applied to make estimations.

The magnitude and the evolution of the RMSD values with the number of ensembles M differed from one matrix potential to another but the difference seemed to be marginal in practice (Figure 6–3). It can be seen from Figure 6–3 that using 30 ensemble members gave stable and satisfactory results using various proportions of the IGBP-Trop dataset as reference data. Using > 30 replicates, we found

practically no change for dataset sizes $N_r = 100, 200, 300, 400$, and 534. The same observation was made in the wet, the intermediate, as well as the dry ranges of the SWRC.

Nemes et al. (2006b) determined that the sufficient minimum number of ensembles for the US National Resources Conservation Service (US NRCS) and the HYPRES datasets were 30 and 50, respectively. They found that when using more than 30 or 50 ensembles, respectively, the effect of adding more ensemble members did not yield any significant changes to the outcome of the estimations, regardless of the reference dataset size. Using the ANN technique, Parasuraman et al. (2006) found also that 30 ensemble members was the optimal number to predict the saturated hydraulic conductivity at the field scale.

Parasuraman et al. (2007b) indicated that adoption of the ensemble technique in the formulation of PTFs helps in addressing one of the pertinent issues in any machine learning algorithm, namely generalization of the estimation results. In our study, 100 replicates were used to generate an ensemble of k-NN estimations. Using this number of replicates can be considered a safe choice to negate the impact of any single replicate on the final estimation results and obtain a high level of generalization of the results.



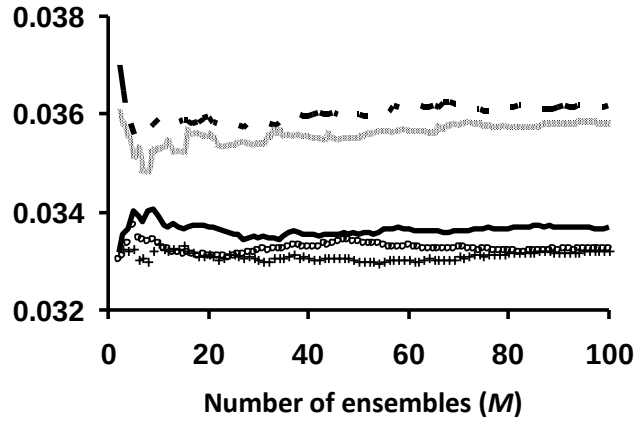


Figure 6–3. Running root mean squared differences (RMSDs) for the Lower Congo test dataset for up to 100 ensembles using sand, silt, clay, bulk density, organic carbon, pH and cation exchange capacity as input attributes and water retention at (a) -1 kPa, (b) -20 kPa and (c) -1500 kPa as output attributes.

6.3.2 Optimizing the k and p terms

A next important preliminary step in establishing the k-NN PTF is the optimization of the two design parameters k and p . A gradual change in both parameters simultaneously enabled us to find an optimal combination of the k and p terms for the given task.

Figure 6–4 shows the interdependence of the k and p terms and N_r , the number of samples in the reference dataset. Estimations developed from smaller data subsets (e.g., here $N_r = 100$ or 200) are more sensitive to changes in k and p . Including more samples from the reference dataset in each individual estimation (i.e., increasing k) beyond a threshold will generally yield worse estimations. This is because with small N_r , an increasing k will mean that a relatively large proportion of the dataset is included in the estimation, rather than a small, but more specific set of samples with very similar characteristics to the target sample. Hence, the estimates will tend to come closer and closer to the reference dataset mean, yielding less accurate ‘local’ estimates. This effect can be further enhanced by the choice of the p (weight) term, as best seen in Figure 6–4a. The closer p is to zero, the more equal the weights are distributed among the chosen k number of samples. When k is relatively large and p is kept small, even less similar samples will have a relatively large weight in the formulation of the final water retention estimate. On the contrary, the effect of a relatively large p value is that even if more samples are used in the individual estimation (i.e., k is increased), the nearest samples (in their properties) would receive a very high proportion of the weights while formulating the final estimate. In essence, a large p value can counteract the potentially negative effect of choosing a k value that is too large. This effect is best seen when k is disproportionally high compared with N_r , as, e.g., in Figure 6–4a.

This combined effect is less and less expressed with an increasing size of the reference dataset (N_r), at least within the examined range of k and p values. It is likely that, following the above logic, with a further increase of k , we would see more impact of the choice of p on the estimation quality when larger N_r values are examined. Nevertheless, p should not be set too high, either; that carries the risk of giving too much weight to one or two individual samples, which may not best represent the characteristics of all similar samples. The simultaneous optimization of the k and p terms requires attentive consideration and good understanding of the underlying effects and consequences.

Based on Figure 6–4, we tried to determine the k number that corresponds to the lowest RMSD (averaged across the eight matric potentials) for p values of 0.5, 1.0, 1.5, 2.0, and 2.5 for dataset sizes N_r of 100, 200, 300, 400, and 534, respectively. An average of all the optimal k numbers determined

for p values of 0.5, 1.0, 1.5, 2.0, and 2.5 was calculated for each reference dataset size (Table 6–1). Because k can only be an integer, the calculated and rounded average k values found in Table 6–1 are plotted in Figure 6–5 against the dataset size. An increasing trend with increasing dataset size was found and the best-fitting equation relating the k number to the reference dataset size N_r was derived based on a power function:

$$k = 0.724 N_r^{0.468} \quad [6-11]$$

Nemes et al. (2006a) found also a power function for the US NRCS dataset (see Equation 6–6). The derived equation (Equation 6–11) yielded values of k very similar to the ones found by Nemes et al. (2006a) for their dataset. Table 6–2 compares the k values derived from the equation of Nemes et al. (2006a) and the ones derived from the Equation 6–11. As noted above, values of k in their study and this study are rounded to the nearest integer, so the actual difference between k values may be even smaller.

To find the best combination of k and p values, we compared the RMSDs provided by each combination of k and p values for each reference dataset size using contour plots (not shown here). The best p value was derived from the intersection between the average k value given in Table 6–1 and the lowest RMSD (three decimal places considered). We did not find a common trend for p values with the reference dataset size; however, for $N_r = 100, 400$ and 534 , we found values around 1. Nemes et al. (2006a) found that the p value ranged from 0.95 to 1.10. For $N_r = 200$ and 300 , the best p values were surprisingly close to 3.0 and 2.2, respectively, which are quite large values. Even if a value of p around 1 were chosen for $N_r = 200$ and 300 , however, the RMSD would increase by only $0.001 \text{ m}^3 \text{ m}^{-3}$; therefore, $p = 1$ seems to be a safe choice. This is in line with the findings and recommendations by Nemes et al. (2006a) regarding the relative insensitivity of the method to a range of p values. Because the difference and its influence appeared to be negligible, we decided to use the function previously used by Nemes et al. (2006a) that relates the p value to the reference dataset size (see Equation 6–7). Hence, a p value of 1.04 corresponds to the full dataset of 534 soil samples of the IGBP-Trop dataset. This value is close to 1, which, following Nemes et al. (2006a), represents a simple inverse relationship between the weight and the distance of the selected sample. The generic settings of the k and p terms that were worked out for temperate-climate soils from the US matched closely with the optimal settings found for the IGBP-Trop dataset. In their study, Patil et al. (2012) also used the functions for k and p provided by Nemes et al. (2006a) and the reference dataset provided with the k-Nearest software (Nemes et al., 2008) and obtained good results for swelling–shrinking soils ($\text{RMSD} < 0.05 \text{ m}^3 \text{ m}^{-3}$).

Table 6-1. Number of nearest neighbours (k) corresponding to the lowest RMSD for different values of p and different dataset sizes N_r .

	$N_r=100$	$N_r=200$	$N_r=300$	$N_r=400$	$N_r=534$
p	k				
0.5	1†	7	9	11	11
1.0	2†	7	10	11	12
1.5	4	7	10	11	13
2.0	7	10	10	13	13
2.5	7	15	14	13	14
Average‡	6	9	11	12	13

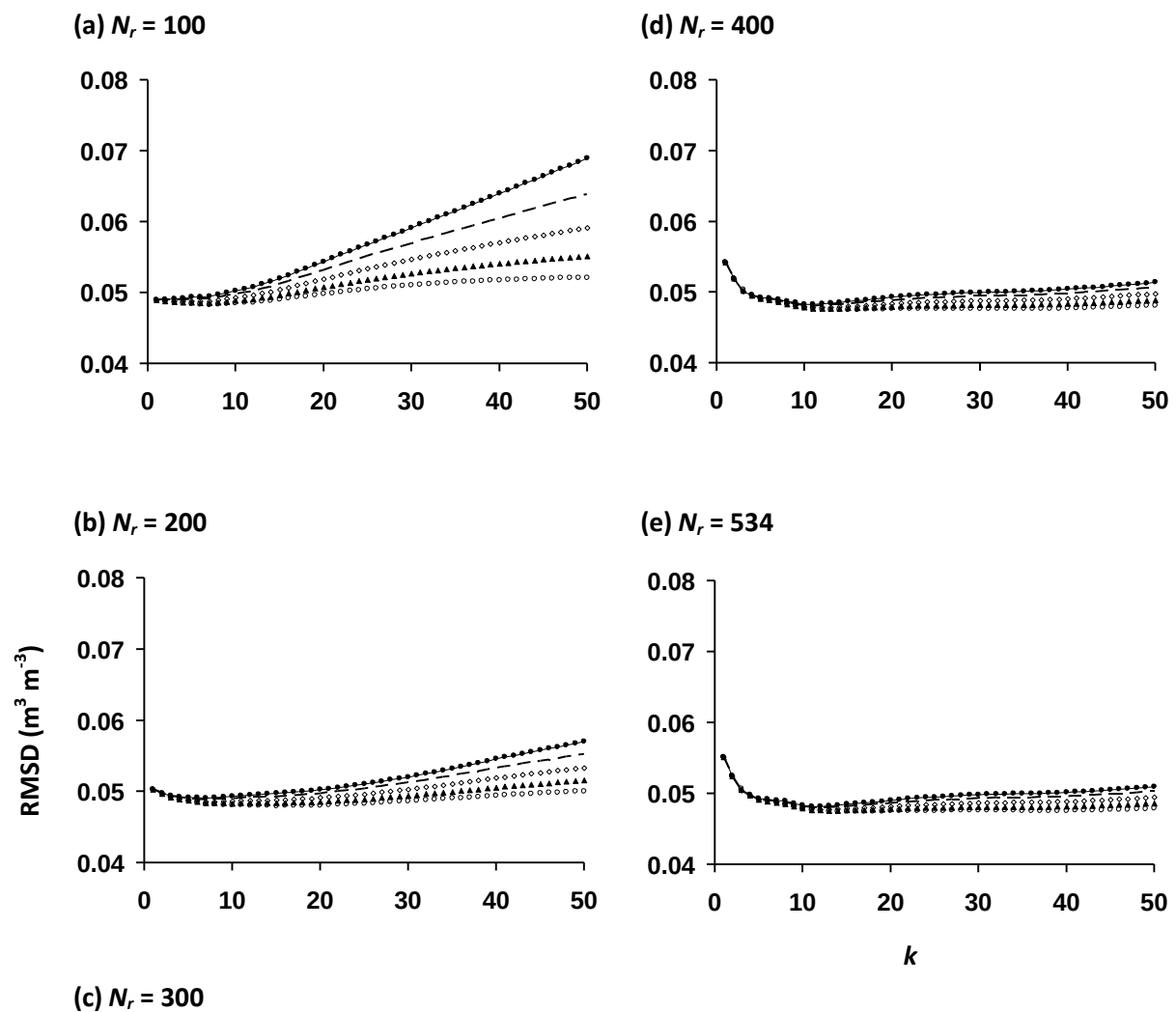
† These values were not taken into account in the calculation of the average k because they did not correspond to a global or a local minimum for RMSD.

‡ Average values are rounded to the nearest integer.

Table 6-2. Comparison of the k number generated by the power function of Nemes et al. (2006a) and the power function derived for this study.†

N_r	k calculated from Nemes et al. (2006a) function	k calculated from the present function
100	6	6
200	9	9
300	11	10
400	13	12
534	14	14

† N_r = reference (training) dataset size.



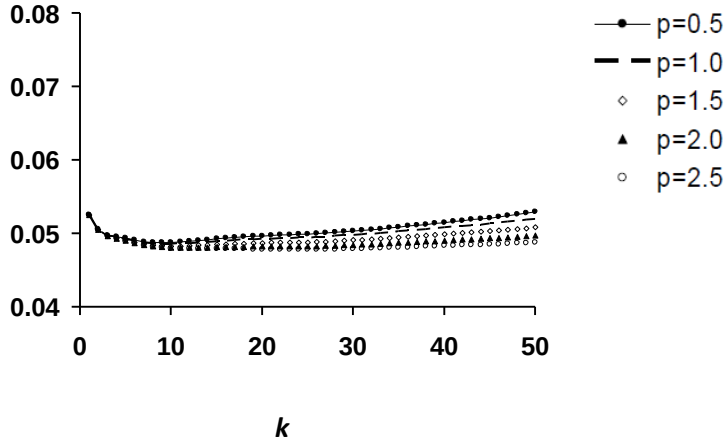


Figure 6-4. Variations of the root mean squared differences (RMSDs) with the number of nearest neighbours k in function of p values and reference dataset sizes N_r .

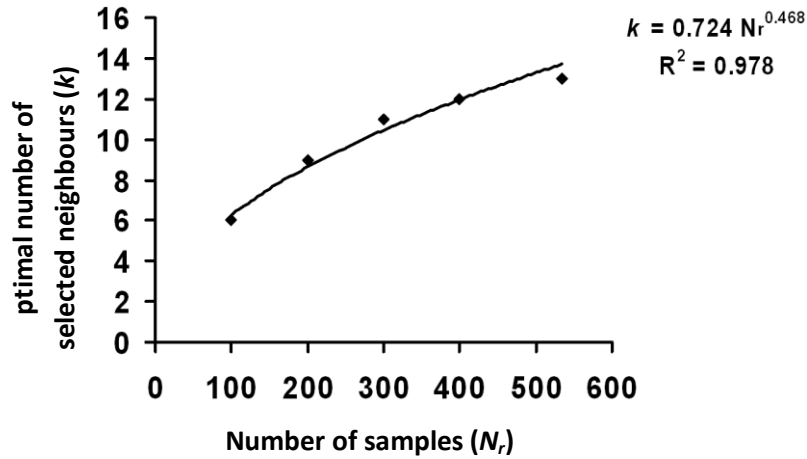


Figure 6-5. Effect of dataset size on the optimal choice of the number of selected neighbours.

6.3.3 Prediction of water retention from an international ‘tropical’ database

In this study, 14 combinations of input soil attributes were used to predict the eight water retention outputs. Table 6-3 gives a summary of the results in terms of MD, RMSD, and R^2 at all eight matric potentials, with the optimized settings and the various combinations of input parameters. The prediction performance of this k-NN algorithm is satisfactory in most cases. When considering individual MD, RMSD, and R^2 values, we found: $-0.009 \text{ m}^3 \text{ m}^{-3} < \text{MD} < 0.055 \text{ m}^3 \text{ m}^{-3}$, $0.032 \text{ m}^3 \text{ m}^{-3} < \text{RMSD} < 0.087 \text{ m}^3 \text{ m}^{-3}$, and $0.280 < R^2 < 0.921$. The average MD, RMSD, and R^2 of eight matric potentials for each input variable combination was: $0.0066 \text{ m}^3 \text{ m}^{-3} < \text{average MD} < 0.0305 \text{ m}^3 \text{ m}^{-3}$, $0.0439 \text{ m}^3 \text{ m}^{-3} < \text{average RMSD} < 0.0619 \text{ m}^3 \text{ m}^{-3}$, and $0.7010 < \text{average } R^2 < 0.8029$. The RMSD values were situated between 0.051 and 0.063 $\text{m}^3 \text{ m}^{-3}$ for prediction of $\theta_{-10\text{kPa}}$ and between 0.032 and 0.038 $\text{m}^3 \text{ m}^{-3}$ for prediction of $\theta_{-1500\text{kPa}}$. These are encouraging results for these two points of the SWRC, which are generally considered to be good approximations of FC and the PWP, respectively, for soils of the humid tropics.

When focusing on the most basic predictor variables—texture (SSC), BD, and OC—generally used in hydraulic PTFs because of their availability in various soil survey reports, it can be seen that the variation in RMSD values is particularly different when BD is or is not included as a predictor (Table

6–3). A marked decreasing trend of RMSD values ($0.076\text{--}0.033\text{ m}^3\text{ m}^{-3}$) from the wet to the dry range of the SWRC can be observed when BD was not considered. On the contrary, when BD was included as predictor, RMSD values were low in the wet range ($< 0.050\text{ m}^3\text{ m}^{-3}$), followed by a slight increase in the intermediate range between matric potentials of -3 and -50 kPa, and a decrease again in the dry range ($< 0.040\text{ m}^3\text{ m}^{-3}$). In the intermediate range, the RMSD yielded by different combinations of input variables varied slightly, with values between 0.050 and $0.060\text{ m}^3\text{ m}^{-3}$; however, the contribution of BD as predictor to the slight decrease in the overall error in prediction in the intermediate range can still be observed. Vereecken et al. (2010) made similar observations regarding the evolution of RMSD values when a combination of SSC, BD, and OM was used as predictors in the published PTFs considered in their review. The matric potentials derived by curve fitting (0 , -50 , and -250 kPa) did not show any out-of-pattern quality in the estimation of water retention.

The results found in this study indicate that the performance of the k-NN algorithm is dependent on the matric potential at which water retention is predicted. Recently, Haghverdi et al. (2012) developed pseudo-continuous ANN PTFs for water retention. Notwithstanding the effect of different combinations of the aforementioned input variables, they also observed relatively large variations in RMSD values as a function of matric potential. For example, the RMSD values were $0.050\text{ m}^3\text{ m}^{-3}$ at -33 kPa and $0.035\text{ m}^3\text{ m}^{-3}$ at -1500 kPa. From Table 6–3 and from previous observations made by several researchers such as Schaap et al. (2001), Vereecken et al. (2010), and Haghverdi et al. (2012), there seems to be an effect of the combinations of different input variables on the quality of the prediction of water contents at various matric potentials. In this study, the difference in prediction performance among models with the 14 input variable combinations was more pronounced in the very wet range of the SWRC (at 0 and -1 kPa), with RMSD values between 0.038 and $0.087\text{ m}^3\text{ m}^{-3}$, and almost negligible in the very dry range of the SWRC (at -250 and -1500 kPa), with RMSD values between 0.034 and $0.040\text{ m}^3\text{ m}^{-3}$. In the intermediate range of the SWRC (from -3 to -50 kPa), the RMSD values yielded by the 14 input combinations were approximately between 0.049 and $0.067\text{ m}^3\text{ m}^{-3}$ (Table 6–3). This can be explained by the major role played by soil structure in the wet and intermediate ranges of the SWRC. Given that the best proxy for soil structure in this study was BD, there was a notable difference in prediction performance between combinations including BD and combinations excluding BD as an input variable.

Table 6–3 shows further that the predictive ability of the k-NN algorithm in terms of bias (MD), overall error (RMSD), and goodness-of-fit (R^2) closely depends on the combination of the ‘predictors’, i.e., the input attributes. Estimation quality may differ significantly when one set of input attributes is used instead of another set. For example, the use of OC and pH were found to considerably reduce the quality of the prediction of water retention in the wet range of the SWRC. When OC and pH are present in the input attributes combination, soils seem to be favored in the training dataset that are quite different from the target soil in their hydraulic behavior in the wet range of the SWRC. On the other hand, OC and pH appeared to have a positive effect on the quality of the prediction in the dry range of the SWRC. Likewise, BD contributed greatly to the improvement of the prediction of water retention in the wet range of the SWRC, while that was not the case in the dry range. Besides soil texture, which plays a major role across the whole range of the SWRC, BD contributes greatly to explaining water retention in the wet range of the SWRC, whereas OC is more influential in the dry range. The k-NN approach is thus able to reflect this physical phenomenon. We found that using the complete set of input attributes, i.e., SSC + BD + OC + pH + CEC was not the best option. As shown in Table 6–3, the best combination appeared to be SSC + BD + CEC with the smallest bias error (average MD = $0.0066\text{ m}^3\text{ m}^{-3}$), the smallest overall error (average RMSD = $0.0439\text{ m}^3\text{ m}^{-3}$), and one of the largest goodness-of-fit values (average $R^2 = 0.8018$), closely followed by the combination SSC

+ BD (average MD = $0.0094 \text{ m}^3 \text{ m}^{-3}$, average RMSD = $0.0444 \text{ m}^3 \text{ m}^{-3}$, and average $R^2 = 0.8029$). On the other hand, the worst combination was found to be SSC + pH, with the largest bias error (average MD = $0.0305 \text{ m}^3 \text{ m}^{-3}$), the largest overall error (average RMSD = $0.0619 \text{ m}^3 \text{ m}^{-3}$), and the smallest goodness-of-fit value (average $R^2 = 0.7010$). One of the reasons for this result could be the lack of a meaningful relationship between pH and water retention at all the matric potentials in the test dataset, with Pearson correlation coefficients $r < 0.203$. Another reason could be the difference in distribution of pH values in the reference and test datasets (Figure 6–2). In the reference dataset, the distribution of pH values was somewhat skewed, whereas in the test dataset, the pH values were normally distributed. This suggests that pH will not be able to provide the information necessary to identify the most similar instances to a given target soil in relation to water retention. The variable pH thus has a limited relationship with water retention and could worsen the prediction of water retention, particularly at high matric potentials, i.e., in the wet range of the SWRC, at least using these particular datasets. Hodnett and Tomasella (2002) found that pH contributed to the estimation of all four parameters of the van Genuchten (1980) equation, possibly because it may be a crude indicator of the degree of weathering of soils in the tropics.

In their study on Vertisols, Patil et al. (2012) found that the inclusion of BD as predictor in the k-NN technique led to a slight increase of the RMSD. They indicated that the BD of Vertisols is known to change with soil water content (swelling-shrinking soils). This particular behaviour was observed and studied by various authors (e.g., Braudeau et al., 2004; Cornelis et al., 2006).

Bulk density and CEC are good indirect indicators of the structure of the soil. Bulk density gives an indication of total soil porosity, whereas CEC gives indications about the clay mineralogy of the soil, which is also responsible for the structural development and porous behaviour of the soil, besides retention of water by adsorption. Pachepsky and Rawls (2003) indicated that BD is a measurable continuous variable that is indirectly related to soil structure. In the same vein, Tranter et al. (2007) proposed a conceptual model that considers BD as the result of particle packing and soil structure. Bronick and Lal (2005) stated that clay minerals influence properties that affect aggregation: surface area, CEC, charge density, dispersivity, and expandability. Based on CEC values, a distinction can be made between soils with high-activity clays and soils with low-activity clays (LAC). Low-activity clays such as kaolinite and halloysite generally occur in highly weathered soils (e.g., Acrisols and Ferralsols), whereas high-activity clays such as montmorillonite are present in swelling–shrinking soils (e.g., Vertisols). As is well known, structure has a non-negligible influence on water retention at high matric potentials. High CEC values are indications of soils with high water retention capacity and poor internal drainage, whereas the opposite is true for soils with low CEC values. Hodnett and Tomasella (2002) found that CEC can be a predictor of the van Genuchten (1980) parameters because it may indicate the effect of mineralogy on the water retention capacity of soils in the tropics.

In this study, the addition of OC seemed not to significantly improve the prediction compared with accounting for texture only. Similarly, Puckett et al. (1985) did not use OM or OC as a predictor to derive water retention PTFs due to its low content in the soil samples from the Lower Coastal Plain in the US. In a study on physical properties and moisture retention characteristics of some ‘tropical’ soils in Nigeria, Lal (1978) did not find any effect of OM or OC on water retention. Zacharias and Wessolek (2007) suggested the exclusion of OM or OC as a predictor in classic PTFs and proposed a new PTF that uses only physical properties such as soil texture and BD. On the contrary, Vereecken et al. (2010) observed that including OM or OC as a predictor in ‘temperate’ PTFs of, e.g., Vereecken et al. (1989), Nemes et al. (2003), and Weynants et al. (2009) led to improved predictions, with the lowest RMSD values in the wet range and the very dry range of the SWRC. This can be explained by

the variability of OM or OC present in ‘temperate’ and ‘tropical’ soils, with soils from temperate areas often having a substantial amount and a wider range of OM. This means that OM or OC can be a suitable predictor of water retention of soils in temperate regions. In contrast, OM or OC content is very low in the humid tropics due to the rapid rate of decomposition under high temperatures and abundant rainfall. Therefore, OM or OC may not have the variability to be an important variable in estimating water retention for soils in the humid tropics.

Furthermore, Table 6–3 shows that the bias error (MD) can contribute, to various extents, to the overall error (RMSD). There is a clear trend to overestimate water retention in the wet and middle ranges of the SWRC, whereas there is a small but almost negligible trend to underestimate water retention in the dry range. The training dataset contained 80% LAC soils (i.e., with $\text{CEC} < 20 \text{ cmol}_c \text{ kg}^{-1} \text{ soil}$) and 20% mixed-activity clay (MIX) soils (i.e., with $\text{CEC} 20\text{--}62 \text{ cmol}_c \text{ kg}^{-1} \text{ soil}$), whereas the test dataset contained $> 95\%$ LAC soils. While LAC soils are dominated by kaolinite and sesquioxides, MIX soils contain other clay minerals such as montmorillonite, which have a relatively higher water retention capacity than kaolinite. Williams et al. (1983) observed that the presence of montmorillonite even in quite small amounts in the soil samples was a discriminating property in relation to water retention. In the evaluation study based on a limited test dataset of soils from Lower Congo reported in Chapter 5, we found that the ‘temperate’ PTFs of Gupta and Larson (1979) largely overestimated the water retention of soils in Lower Congo. They attributed this result to the differences in soil properties and in the mineralogy between the test dataset and the dataset used to develop the PTFs. One possible explanation for the large positive bias could be the difference in the distribution of texture classes, with a strong presence of silty soils in temperate (development) soil datasets, while clayey soils dominate in tropical (test) soil datasets. Another reason may be the presence of montmorillonitic soils in the development dataset used by Gupta and Larson (1979) and the large dominance of kaolinitic soils in the independent test dataset used in Chapter 5. In their study of the performance of various PTFs applied to Ferralsols from Cuba, Medina et al. (2002) indicated that clay type plays a vital role in the retention and transmission properties of a given soil. It is the reason why soils in the humid tropics can have much more clay than soils in the temperate regions but a much lower water retention capacity.

Table 6-3. Summary of results in terms of MD, RMSD and R^2 , for the k-NN method with optimized settings at eight different matric potentials and using 14 combinations of input attributes.[†]

Input attributes	Predicted water content								AvgMD
	θ_{0kPa}	θ_{-1kPa}	θ_{-3kPa}	θ_{-10kPa}	θ_{-20kPa}	θ_{-50kPa}	$\theta_{-250kPa}$	$\theta_{-1500kPa}$	
	MD ($m^3 m^{-3}$)								
SSC	0.039 (0.0025)	0.037 (0.0027)	0.013 (0.0033)	0.013 (0.0021)	0.029 (0.0023)	0.032 (0.0022)	-0.004 (0.0018)	-0.005 (0.0016)	0.0193
SSC + BD	0.013 (0.0018)	0.014 (0.0020)	-0.006 (0.0031)	0.003 (0.0022)	0.022 (0.0022)	0.027 (0.0021)	0.002 (0.0015)	0.000 (0.0014)	0.0094
SSC + OC	0.048 (0.0026)	0.043 (0.0026)	0.017 (0.0030)	0.016 (0.0027)	0.032 (0.0027)	0.035 (0.0025)	-0.005 (0.0018)	-0.006 (0.0017)	0.0225
SSC + BD + OC	0.021 (0.0021)	0.019 (0.0022)	-0.002 (0.0030)	0.006 (0.0024)	0.024 (0.0023)	0.028 (0.0022)	-0.001 (0.0017)	-0.003 (0.0016)	0.0115
SSC + pH	0.053 (0.0031)	0.052 (0.0031)	0.021 (0.0037)	0.029 (0.0033)	0.044 (0.0032)	0.046 (0.0030)	-0.001 (0.0017)	0.000 (0.0016)	0.0305
SSC + CEC	0.038 (0.0027)	0.036 (0.0028)	0.011 (0.0030)	0.011 (0.0025)	0.027 (0.0026)	0.030 (0.0024)	-0.005 (0.0017)	-0.004 (0.0017)	0.0180
SSC + pH + CEC	0.049 (0.0027)	0.048 (0.0026)	0.017 (0.0030)	0.026 (0.0029)	0.042 (0.0028)	0.044 (0.0026)	0.000 (0.0018)	0.001 (0.0017)	0.0284
SSC + OC + pH	0.055 (0.0023)	0.051 (0.0022)	0.019 (0.0027)	0.027 (0.0026)	0.042 (0.0026)	0.044 (0.0024)	-0.001 (0.0020)	0.000 (0.0019)	0.0296
SSC + OC + CEC	0.049 (0.0029)	0.044 (0.0028)	0.016 (0.0029)	0.014 (0.0028)	0.029 (0.0028)	0.032 (0.0026)	-0.008 (0.0018)	-0.008 (0.0017)	0.0210
SSC + BD + pH	0.017 (0.0018)	0.018 (0.0018)	-0.005 (0.0027)	0.012 (0.0026)	0.029 (0.0026)	0.034 (0.0024)	0.002 (0.0020)	0.004 (0.0019)	0.0139
SSC + BD + CEC	0.012 (0.0018)	0.013 (0.0020)	-0.009 (0.0030)	-0.001 (0.0021)	0.018 (0.0021)	0.023 (0.0020)	-0.002 (0.0017)	-0.001 (0.0016)	0.0066
SSC + BD + pH + CEC	0.016 (0.0018)	0.018 (0.0019)	-0.005 (0.0028)	0.011 (0.0026)	0.029 (0.0026)	0.033 (0.0025)	0.002 (0.0022)	0.003 (0.0020)	0.0134
SSC + OC + pH + CEC	0.053 (0.0025)	0.050 (0.0023)	0.018 (0.0026)	0.026 (0.0026)	0.041 (0.0026)	0.044 (0.0024)	-0.001 (0.0020)	0.000 (0.0019)	0.0289
SSC + BD + OC + pH + CEC	0.023 (0.0021)	0.023 (0.0020)	-0.001 (0.0026)	0.015 (0.0025)	0.032 (0.0025)	0.036 (0.0023)	0.000 (0.0021)	0.001 (0.0019)	0.0161

[†] Standard deviations of MD, RMSD and R^2 values generated by ensemble of k-NN estimations based on 100 replicates are presented in parentheses. SSC is sand (%), silt (%) and clay (%), BD is bulk density ($Mg m^{-3}$), OC is organic carbon (%), pH is acidity (-), CEC is cation exchange capacity ($cmol_c kg^{-1}$ soil).

Table 6–3. Summary of results in terms of MD, RMSD and R^2 , for the k-NN method with optimized settings at eight different matric potentials and using 14 combinations of input attributes (continued).†

Input attributes	Predicted water content								AvgRMSD
	θ_{0kPa}	θ_{-1kPa}	θ_{-3kPa}	θ_{-10kPa}	θ_{-20kPa}	θ_{-50kPa}	$\theta_{-250kPa}$	$\theta_{-1500kPa}$	
	RMSD ($m^3 m^{-3}$)								
SSC	0.076 (0.0017)	0.070 (0.0017)	0.060 (0.0022)	0.057 (0.0016)	0.055 (0.0019)	0.054 (0.0019)	0.037 (0.0009)	0.035 (0.0008)	0.0555
SSC + BD	0.039 (0.0014)	0.039 (0.0014)	0.050 (0.0020)	0.051 (0.0013)	0.050 (0.0016)	0.051 (0.0017)	0.039 (0.0010)	0.036 (0.0008)	0.0444
SSC + OC	0.078 (0.0022)	0.070 (0.0021)	0.057 (0.0021)	0.058 (0.0018)	0.057 (0.0024)	0.056 (0.0023)	0.034 (0.0012)	0.033 (0.0013)	0.0554
SSC + BD + OC	0.048 (0.0024)	0.045 (0.0021)	0.051 (0.0023)	0.052 (0.0016)	0.050 (0.0017)	0.050 (0.0017)	0.034 (0.0010)	0.033 (0.0010)	0.0454
SSC + pH	0.086 (0.0023)	0.078 (0.0021)	0.058 (0.0020)	0.063 (0.0026)	0.067 (0.0033)	0.067 (0.0032)	0.039 (0.0007)	0.037 (0.0007)	0.0619
SSC + CEC	0.075 (0.0023)	0.071 (0.0022)	0.060 (0.0024)	0.057 (0.0016)	0.056 (0.0020)	0.055 (0.0020)	0.040 (0.0010)	0.038 (0.0009)	0.0565
SSC + pH + CEC	0.083 (0.0022)	0.076 (0.0020)	0.058 (0.0021)	0.060 (0.0019)	0.063 (0.0025)	0.063 (0.0024)	0.039 (0.0008)	0.038 (0.0009)	0.0600
SSC + OC + pH	0.087 (0.0019)	0.077 (0.0017)	0.058 (0.0019)	0.062 (0.0019)	0.064 (0.0023)	0.063 (0.0022)	0.037 (0.0012)	0.035 (0.0013)	0.0604
SSC + OC + CEC	0.080 (0.0021)	0.072 (0.0019)	0.058 (0.0020)	0.058 (0.0017)	0.056 (0.0024)	0.054 (0.0023)	0.035 (0.0010)	0.033 (0.0011)	0.0558
SSC + BD + pH	0.045 (0.0016)	0.042 (0.0015)	0.049 (0.0020)	0.053 (0.0012)	0.055 (0.0018)	0.057 (0.0018)	0.038 (0.0008)	0.036 (0.0007)	0.0469
SSC + BD + CEC	0.038 (0.0014)	0.038 (0.0014)	0.049 (0.0020)	0.051 (0.0012)	0.050 (0.0015)	0.050 (0.0016)	0.039 (0.0009)	0.036 (0.0007)	0.0439
SSC + BD + pH + CEC	0.045 (0.0017)	0.042 (0.0016)	0.049 (0.0020)	0.053 (0.0012)	0.055 (0.0018)	0.056 (0.0018)	0.038 (0.0007)	0.036 (0.0007)	0.0468
SSC + OC + pH + CEC	0.086 (0.0021)	0.077 (0.0019)	0.058 (0.0020)	0.061 (0.0018)	0.063 (0.0023)	0.062 (0.0022)	0.036 (0.0011)	0.035 (0.0012)	0.0598
SSC + BD + OC + pH + CEC	0.052 (0.0025)	0.047 (0.0022)	0.051 (0.0023)	0.054 (0.0015)	0.056 (0.0019)	0.056 (0.0019)	0.035 (0.0009)	0.032 (0.0009)	0.0479

† Standard deviations of MD, RMSD and R^2 values generated by ensemble of k-NN estimations based on 100 replicates are presented in parentheses. SSC is sand (%), silt (%) and clay (%), BD is bulk density ($Mg m^{-3}$), OC is organic carbon (%), pH is acidity (-), CEC is cation exchange capacity ($cmol_c kg^{-1}$ soil).

Table 6–3. Summary of results in terms of MD, RMSD and R^2 , for the k-NN method with optimized settings at eight different matric potentials and using 14 combinations of input attributes (continued).[†]

Input attributes	Predicted water content								AvgR ²
	θ_{0kPa}	θ_{-1kPa}	θ_{-3kPa}	θ_{-10kPa}	θ_{-20kPa}	θ_{-50kPa}	$\theta_{-250kPa}$	$\theta_{-1500kPa}$	
	R^2 (–)								
SSC	0.315 (0.0225)	0.400 (0.0216)	0.604 (0.0144)	0.844 (0.0094)	0.888 (0.0086)	0.894 (0.0079)	0.910 (0.0040)	0.910 (0.0039)	0.7206
SSC + BD	0.661 (0.0140)	0.654 (0.0114)	0.654 (0.0089)	0.851 (0.0057)	0.891 (0.0058)	0.893 (0.0056)	0.908 (0.0038)	0.911 (0.0033)	0.8029
SSC + OC	0.418 (0.0244)	0.496 (0.0236)	0.652 (0.0123)	0.832 (0.0114)	0.879 (0.0112)	0.889 (0.0099)	0.919 (0.0058)	0.917 (0.0066)	0.7503
SSC + BD + OC	0.624 (0.0185)	0.641 (0.0156)	0.667 (0.0090)	0.851 (0.0065)	0.896 (0.0057)	0.901 (0.0054)	0.921 (0.0043)	0.920 (0.0046)	0.8026
SSC + pH	0.280 (0.0201)	0.412 (0.0192)	0.656 (0.0129)	0.791 (0.0160)	0.841 (0.0146)	0.853 (0.0132)	0.888 (0.0041)	0.887 (0.0042)	0.7010
SSC + CEC	0.324 (0.0256)	0.407 (0.0241)	0.619 (0.0123)	0.838 (0.0098)	0.877 (0.0095)	0.883 (0.0087)	0.895 (0.0050)	0.898 (0.0049)	0.7176
SSC + pH + CEC	0.311 (0.0202)	0.445 (0.0187)	0.682 (0.0120)	0.812 (0.0109)	0.860 (0.0093)	0.869 (0.0084)	0.886 (0.0045)	0.886 (0.0050)	0.7189
SSC + OC + pH	0.337 (0.0230)	0.452 (0.0200)	0.669 (0.0110)	0.802 (0.0103)	0.859 (0.0088)	0.872 (0.0082)	0.900 (0.0064)	0.898 (0.0077)	0.7236
SSC + OC + CEC	0.404 (0.0233)	0.479 (0.0205)	0.652 (0.0100)	0.827 (0.0113)	0.872 (0.0113)	0.884 (0.0100)	0.917 (0.0050)	0.917 (0.0059)	0.7440
SSC + BD + pH	0.590 (0.0128)	0.644 (0.0105)	0.666 (0.0116)	0.821 (0.0066)	0.863 (0.0062)	0.868 (0.0060)	0.892 (0.0043)	0.898 (0.0038)	0.7803
SSC + BD + CEC	0.676 (0.0131)	0.672 (0.0101)	0.656 (0.0090)	0.841 (0.0067)	0.878 (0.0067)	0.881 (0.0066)	0.902 (0.0043)	0.908 (0.0034)	0.8018
SSC + BD + pH + CEC	0.591 (0.0126)	0.649 (0.0102)	0.670 (0.0113)	0.821 (0.0065)	0.863 (0.0061)	0.867 (0.0057)	0.892 (0.0041)	0.899 (0.0036)	0.7815
SSC + OC + pH + CEC	0.346 (0.0239)	0.461 (0.0201)	0.676 (0.0103)	0.807 (0.0098)	0.864 (0.0084)	0.876 (0.0076)	0.901 (0.0059)	0.899 (0.0071)	0.7288
SSC + BD + OC + pH + CEC	0.574 (0.0174)	0.636 (0.0156)	0.673 (0.0104)	0.824 (0.0076)	0.872 (0.0069)	0.879 (0.0069)	0.910 (0.0049)	0.915 (0.0047)	0.7854

[†] Standard deviations of MD, RMSD and R^2 values generated by ensemble of k-NN estimations based on 100 replicates are presented in parentheses. SSC is sand (%), silt (%) and clay (%), BD is bulk density (Mg m^{-3}), OC is organic carbon (%), pH is acidity (–), CEC is cation exchange capacity ($\text{cmol}_c \text{ kg}^{-1}$ soil).

6.3.4 Prediction performance of the k-nearest neighbour and multiple linear regression approaches

The MLR PTFs of Hodnett and Tomasella (2002) used texture, BD, pH, OC, and CEC as predictors to estimate the van Genuchten (1980) parameters, whereas the point-based PTFs of Minasny and Hartemink (2011) used texture, BD, and OC as inputs. The RMSDs of these PTFs were compared with the k-NN algorithm using different combinations of predictors: SSC + OC, SSC + BD, SSC + BD + CEC, as well as the full set of available predictors (SSC + BD + OC + pH + CEC) (Table 6–4).

An independent one-sample t-test (W.S. Gosset, ‘The probable error of a mean’. Collected papers, p.12) was run, evaluated at the 0.05 significance level, which indicated that the RMSD values generated by the MLR PTFs and the k-NN models were statistically different at each matric potential. The RMSDs of the k-NN models varied by matric potential and which set of predictors was used, but the PTFs of Hodnett and Tomasella (2002) yielded comparable RMSD values to those of the k-NN algorithm with certain combinations of inputs, primarily the SSC + BD and SSC + BD + CEC models. The differences were rather small in most cases, but they were significant in all cases, given the very small standard deviation of ensemble RMSDs. At near saturation, the k-NN estimates were more accurate, but in the intermediate matric potential range (from -10 to -50 kPa) the Hodnett and Tomasella (2002) PTFs yielded smaller RMSD values than the k-NN algorithm. The Hodnett and Tomasella (2002) PTFs and k-NN showed particularly comparable performance in the dry range. We note that one of the points in the intermediate range (i.e., -50 kPa) was derived by curve fitting for the Lower Congo dataset, which may have introduced some degree of extra uncertainty into the estimations. The point-based PTFs of Minasny and Hartemink (2011) gave significantly greater RMSD values than the PTFs of Hodnett and Tomasella (2002) and any of the examined k-NN algorithms at the two available matric potentials (Table 6–4).

Any direct comparison of the performance of PTFs that do not use the same inputs is influenced by the cost and benefit of any extra variable(s), so conclusions have to be drawn carefully. The k-NN algorithm that uses SSC + BD + OC + pH + CEC requires the same input attributes as the PTF of Hodnett and Tomasella (2002) that predicts the van Genuchten (1980) parameters. On the other hand, the k-NN algorithm using SSC + BD uses the same inputs as the -10 kPa PTF of Minasny and Hartemink (2011), while the k-NN algorithm using SSC + OC uses the same inputs as the -1500 kPa PTF of Minasny and Hartemink (2011). In our comparison with the two MLR models, we concluded that the presented k-NN models that use the same inputs showed better performance measures than the Minasny and Hartemink (2011) PTFs. On the other hand, when, e.g., the SSC + BD k-NN model was compared with the Hodnett and Tomasella (2002) PTFs, a somewhat weaker performance was achieved but with a significantly smaller number of inputs—i.e., k-NN did not use OC, pH, or CEC as inputs. It is of particular value in data- and resource-poor environments when the need for input data is minimized in a quest to obtain estimates of expensive but important soil hydraulic properties. The Hodnett and Tomasella (2002) PTFs require the user to have all five of the listed properties available to estimate the water retention of a tropical soil, which can be a serious limitation in their applicability. Our k-NN approach can be used in a hierarchical way, adjusting the used inputs to their availability, and acceptably good and stable estimation results can already be achieved by using only texture and BD as predictors. Among the examined PTFs, our k-NN based PTFs appear to show the best value when statistical performance is combined with the PTFs’ need for input. Given that the source of the development data was the same for the two MLR and the k-NN PTFs, it is likely that the PTF development methodology and the data on which they have been tested are the combined reason for that finding. Given its capability and flexibility in utilizing a limited or a wider range of predictors

hierarchically, based on their availability, the k-NN technique presents a far greater number of choices and flexibility to the user than do published MLR PTFs. Additionally, given that all calculations are made in real time in k-NN, as the expected growth and development of tropical soil databases occurs, those new data can be taken into account by the k-NN technique without the need to redevelop any equations, which would be necessary with MLR PTFs such as those of Hodnett and Tomasella (2002) and Minasny and Hartemink (2011).

In preparation for future needs and increased computing capabilities, the k-NN technique can also readily provide an estimate of the uncertainty when ensembles of estimations are generated. Such advances can be well taken into account while parameterizing simulation-based environmental risk-assessment and scenario studies. The k-NN application also demonstrated how any number of points can be estimated simultaneously on the SWRC curve, given that those points exist in the source database. Therefore, besides its capability to provide SWRC estimates of competitive quality, the proposed k-NN approach gives a number of additional benefits to the user compared with existing MLR approaches. When provided with an enhanced user interface, similar in nature to the k-Nearest software of Nemes et al. (2008), the k-NN variant developed in this study can be easily implemented by potential users interested in soils of the humid tropics.

Table 6-4. Prediction performance in terms of RMSD of the k-NN method using four combinations of input attributes, of the PTFs of Hodnett and Tomasella (2002) and Minasny and Hartemink (2011).[†]

PTFs	RMSD (m ³ m ⁻³)							
	θ_{0kPa}	θ_{-1kPa}	θ_{-3kPa}	θ_{-10kPa}	θ_{-20kPa}	θ_{-50kPa}	$\theta_{-250kPa}$	$\theta_{-1500kPa}$
k-NN (SSC+BD)	0.039 (0.0014)	0.039 (0.0014)	0.050 (0.0020)	0.051 (0.0013)	0.050 (0.0016)	0.051 (0.0017)	0.039 (0.0010)	0.036 (0.0008)
k-NN (SSC+OC)	0.078 (0.0022)	0.070 (0.0021)	0.057 (0.0021)	0.058 (0.0018)	0.057 (0.0024)	0.056 (0.0023)	0.034 (0.0012)	0.033 (0.0013)
k-NN (SSC+BD+CEC)	0.038 (0.0014)	0.038 (0.0014)	0.049 (0.0020)	0.051 (0.0012)	0.050 (0.0015)	0.050 (0.0016)	0.039 (0.0009)	0.036 (0.0007)
k-NN (SSC+BD+OC+pH+CEC)	0.052 (0.0025)	0.047 (0.0022)	0.051 (0.0023)	0.054 (0.0015)	0.056 (0.0019)	0.056 (0.0019)	0.035 (0.0009)	0.032 (0.0009)
Hodnett and Tomasella (2002)	0.036	0.042	0.059	0.049	0.046	0.041	0.036	0.035
Minasny and Hartemink (2011)	-	-	-	0.062	-	-	-	0.045

[†] Standard deviations of RMSD values generated by ensemble of k-NN estimations based on 100 replicates are presented in parentheses. SSC is sand (%), silt (%) and clay (%), BD is bulk density (Mg m⁻³), OC is organic carbon (%), pH is acidity (-), CEC is cation exchange capacity (cmol_c kg⁻¹ soil).

6.4 Conclusions

A variant of the k-NN algorithm developed by Nemes et al. (2006a) was applied and tested to predict the water retention of soils from Lower Congo in Central Africa based on an international dataset (IGBP-Trop) of soils of the (sub)humid tropics. Two design parameters, k and p , that are user defined and determined before and independent of applying the nonparametric k-NN algorithm were optimized to take better advantage of the k-NN variant introduced in this study. The optimized k and p values were found to be similar to those of previous studies. The results showed that this k-NN variant was able to estimate water retention at eight different matric potentials (0, -1, -3, -10, -20, -50, -250, and -1500 kPa), i.e., from the wet to the dry range of the SWRC, with an average RMSD < 0.046 m³ m⁻³ when SSC + BD or SSC + BD + CEC were selected as input variables. The overall prediction performance of the proposed nonparametric approach was compared with two tropical equation-based PTFs of Hodnett and Tomasella (2002) and Minasny and Hartemink (2011) based on the MLR approach. The results suggest that the k-NN approach shows comparable prediction performance to the examined MLR PTFs, which makes it a competitive alternative to those equation-based PTFs that are currently available to predict the water retention of soils in the humid tropics.

While performing similarly, the presented k-NN variant provides a great degree of flexibility and extra options to the user. The user can, for example, (1) incorporate additional data by appending to or replacing the reference database without the need for or burden of developing new equations, (2) develop the estimations in real time, decide in real time what inputs to use, and vary them from sample to sample if desired, (3) estimate any number and combination of SWRC points simultaneously, driven by their availability in the reference or development dataset, and (4) generate an uncertainty measure for the estimates. These advantages can be particularly beneficial in the context of developing countries, where there is growing demand—as well as potential—to continuously develop soil databases and subsequent simulation-based studies for pedological, agricultural, and environmental studies. For future research, we recommend testing the ability of this technique to predict the water retention of other soils found in the tropics, for example volcanic soils that present some specific properties. These soils present a completely different mineralogy than highly weathered soils or swelling–shrinking soils and may need a completely different reference or training dataset than the IGBP-Trop dataset to provide acceptable estimations of their hydraulic characteristics.

7 Hierarchical pedotransfer functions to predict bulk density of Lower Congo soils

7.1 Introduction

Soil bulk density (BD) is defined as the mass of an oven-dry sample of undisturbed soil per unit bulk volume (ISSS Working Group, 1998, p.153). Assouline (2006) wrote that an increase in BD influences many aspects of the soil–water–plant–atmosphere system. Bulk density is a key soil property for sustainable land management in various agro-climatic zones in the world. The influence of many soil management practices on soil properties that affect water movement and water retention in soils can be reflected in the soil BD (Rawls et al., 1983).

It is considered as a key indicator of soil structure and can reflect the compaction and the porosity of a soil (Tamminen and Starr, 1994). Bulk density is also used as one of the predictors of the water retention characteristics of various soils in temperate as well as in tropical regions. In different studies related to soils, BD is a critical characteristic that is used as one of the main inputs in various agricultural and environmental models.

In the effort to mitigate climate change through accumulation of carbon in the soil, national estimates of soil carbon stocks are increasingly demanded (Eswaran et al., 1993). To respond to this need, Schwartz and Namri (2002) mapped the total organic carbon in the soils of the Republic of the Congo (R. Congo). Soil organic carbon (OC) estimates require data on BD which serve as a conversion factor. Indeed, the BD values are needed to convert carbon contents as percent of dry weight to weight of carbon per unit area (Howard et al., 1995; Arrouays et al., 2001). Since ‘tropical’ soils represent a great part of the global carbon stocks, it is important to determine the carbon storage in such soils (Schwartz and Namri, 2002). Recently, Hollis et al. (2012) wrote that the assessment of soil carbon and nutrient stocks or environmental risks is required for implementation of various environmental policies in Europe.

However, data on BD are generally missing in various soil survey reports in tropical countries particularly in Central Africa (Schwartz and Namri, 2002; Batjes, 2008). In their attempt to estimate carbon stocks in R. Congo soils, Schwartz and Namri (2002) used approximately 700 soil profiles with data on OC content whereas BD data were available for only 33 profiles, all located in the southern part of the country. The paucity of BD data is attributed to the fact that its measurements are time-consuming, labour-intensive, expensive and often impractical for environmental studies on large areas (Bernoux et al., 1998). Therefore, pedotransfer functions (PTFs) have been developed to estimate BD based on data routinely measured or registered in soil databases or easily acquired (Bouma, 1989), the main premise being the characterization of dependence structures between the response variable (BD) and other predictor variables (e.g., texture, organic matter OM or OC) (Vereecken and Herbst, 2004).

Stewart et al. (1970) and Adams (1973) indicated that the main factors determining BD of a soil layer are its OM content (and its associated density) and its mineral density as determined by its texture, either the sand or clay content or both. Thus, OC, clay and/or sand content are generally considered by several authors as the main predictors of BD and are present in various PTFs based on regression analysis (Huntington et al., 1989; Manrique and Jones, 1991; Bernoux et al., 1998; Tomasella and Hodnett, 1998; Kaur et al., 2002; Benites et al., 2007; Hollis et al., 2012). In various BD PTFs developed for soils of temperate climates, an organic variable notably OC (e.g., Alexander, 1980; Manrique and Jones, 1991), OM (e.g., Federer, 1983; Huntington et al., 1989) or loss-on-ignition (designated as LOI, an easy and low-cost method for determination of OM used by, e.g., Jeffrey, 1970; Adams, 1973; Harrison and Bockock, 1981; Honeysett and Ratkowsky, 1989; Tamminen and Starr, 1994; De Vos et al., 2005) has been presented as the sole predictor of BD. This is mainly because of its major role as a binding agent of soil aggregates and as a nucleus in the formation of aggregates

(Bronick and Lal, 2005), particularly in forest soils rich in OM (De Vos et al., 2005). However, Calhoun et al. (2001) indicated that OC and texture present apparent limitations in their predictive ability as they can only explain between 50 and 60% of variation in BD. Therefore, Tranter et al. (2007) perceived the need to improve the prediction of BD by two main approaches: (1) the inclusion of better predictors and (2) the development of improved modelling methods.

On the other hand, many authors studied and emphasized the major role of Fe and Al oxides in the structural characteristics of soils related to their aggregative effect on soil particles (Schwertmann, 1988a,b; Chesters et al., 1957; Arshad et al., 1980; Banejee et al., 1981; Krishna Murti and Huang, 1987; Arduino et al., 1989; Goldberg et al., 1990; Colombo and Torrent, 1991; Torrent, 1994). Due to high temperature and abundant rainfall, the OM present in the soil under tropical climates decomposes much more rapidly compared to the OM present in soils of temperate climates. This partly explains the low OC content of highly weathered soils located in the humid tropics and particularly under savannah vegetation. In these soils, Fe and Al oxides are known to play the major role in the binding and the stability of the aggregates. Fe and Al oxides were found to affect other soil properties such as cation exchange, anion adsorption, moisture retention and colour (Schwertmann, 1988a; Schwertmann and Taylor, 1989). Because of their great effect on cation and anion exchange capacity, on coloration and chemical fertility, on aggregation and cementation phenomena of various soils, several authors studied various properties and phenomena related to Fe and Al oxides in soils (Mekaru and Uehara, 1972; Blackmore, 1973; Bigham et al., 1978; Schwertmann et al., 1979; Arshad et al., 1980; Banejee et al., 1981; Borggaard, 1983; Schulze and Schwertmann, 1984; Schwertmann and Kampf, 1985; Schwertmann and Latham, 1986; Anand and Gilkes, 1987; Schwertmann, 1988a,b; Schwertmann and Taylor, 1989; Arduino et al., 1989; Goldberg et al., 1990; Colombo and Torrent, 1991; Friedl and Schwertmann, 1996; Van Ranst et al., 1998; Baert and Van Ranst, 1998; Baert et al., 1999).

As the influence of Fe and Al oxides on soil structure as the major binding agents of aggregates is widely recognized for soils of the humid tropics, there is an interest to investigate their potential as predictors of BD particularly for highly weathered soils. To our knowledge, the few PTFs which used dithionite-citrate-bicarbonate extractable iron (DCB-Fe) and aluminium (DCB-Al) as predictors of BD in the humid tropics have been developed by Bernoux et al. (1998) and Benites et al. (2007) for soils from Brazil. In their studies, clay, OC, DCB-Fe and DCB-Al were found to be good predictors of BD of highly weathered soils (e.g., Ferralsols and Acrisols). These soils were predominant in their dataset and showed high aggregate stability in which aggregates of clay are stabilized by high contents of Fe oxides and by OM. However, Bernoux et al. (1998) and Benites et al. (2007) did not put much emphasis on DCB-Fe and DCB-Al as potential predictors of BD. They argued that data on Fe and Al content are not commonly available in Brazilian soil survey reports. Therefore, more studies should be performed in other countries/regions of the humid tropics in order to confirm the relative importance of DCB-Fe and DCB-Al as predictors of BD of strongly weathered soils. For D.R. Congo and particularly for the Lower Congo region, there is a great opportunity to conduct such a study because a huge quantity of physico-chemical data among which texture (sand, silt, clay), OC, DCB-Fe and DCB-Al have been acquired and are therefore available in various soil survey reports (e.g., Sys, 1972; Baert et al., 1991a; Baert et al., 1993; Baert et al., 2009).

On the contrary, BD data are often missing in most soil datasets of Central Africa and this limits the usefulness of the past soil survey reports or databases for global estimation of carbon pools in the Congo basin. To circumvent this limitation, BD needs to be estimated from other available soil properties by means of PTFs. Therefore, the objectives of this study are: (1) to compare the predictive capability of some published ‘tropical’ and ‘temperate’ BD PTFs on the Lower Congo dataset; (2) to

formulate hierarchical PTFs (i.e. PTFs with limited to more extended sets of predictors) developed based on a dataset of undisturbed soils to predict BD of soils in Lower Congo with two different approaches: an equation-based approach and a pattern-recognition approach; (3) to investigate the potential of DCB-Fe and DCB-Al as predictors of BD of highly weathered soils of Central Africa.

7.2 Materials and Methods

7.2.1 Study area: Lower Congo

The study area has been described in Chapter 2.

7.2.2 Lower Congo soil dataset

A dataset of 196 undisturbed soil samples was used as a training and test dataset to develop BD PTFs for Lower Congo soils. These soils are mainly highly weathered soils classified according to IUSS Working Group WRB (2006) as Ferralsols, Acrisols, Nitisols and Arenosols (sandy soils). One of the consequences of the advanced weathering is that most of these soils, except the Arenosols, have high clay content. The clay fraction consists mainly of low activity clays: kaolinite and Fe oxides (goethite and hematite).

Soil samples were collected under different land uses (savannah, forest, agricultural fields and in old quarries). Bulk density was determined by the core method following the procedures of Grossman and Reinsch (2002). After moistening the samples up to saturation, they were gradually subjected to a tension of -10 kPa and subsequently weighted. Water content was then determined on a representative subsample and used to determine the sample's oven-dry (105°C) weight as detailed in Cornelis et al. (2005). It should be noted that none of the samples was subjected to shrinkage upon drying.

The physico-chemical characteristics of all soil samples (fine earth) taken from the cores were determined using standard methods described in detail in Van Ranst et al. (1999). During these analyses, texture (by the pipette method of Köhn, 1929), OC, pH, CEC, DCB-Fe and DCB-Al were determined. Selected soil properties used to develop BD PTFs were the following: sand ($50\text{--}2000\text{ }\mu\text{m}$), silt ($2\text{--}50\text{ }\mu\text{m}$), and clay content ($< 2\text{ }\mu\text{m}$) according to the USDA classification system (Soil Survey Staff, 1951), OC, pH, CEC, DCB-Fe and DCB-Al. The textural distribution of the Lower Congo soil dataset is shown in Figure 7-1. Box plots of the selected soil attributes, for the training/test dataset of soils from Lower Congo are given in Figure 7-2.

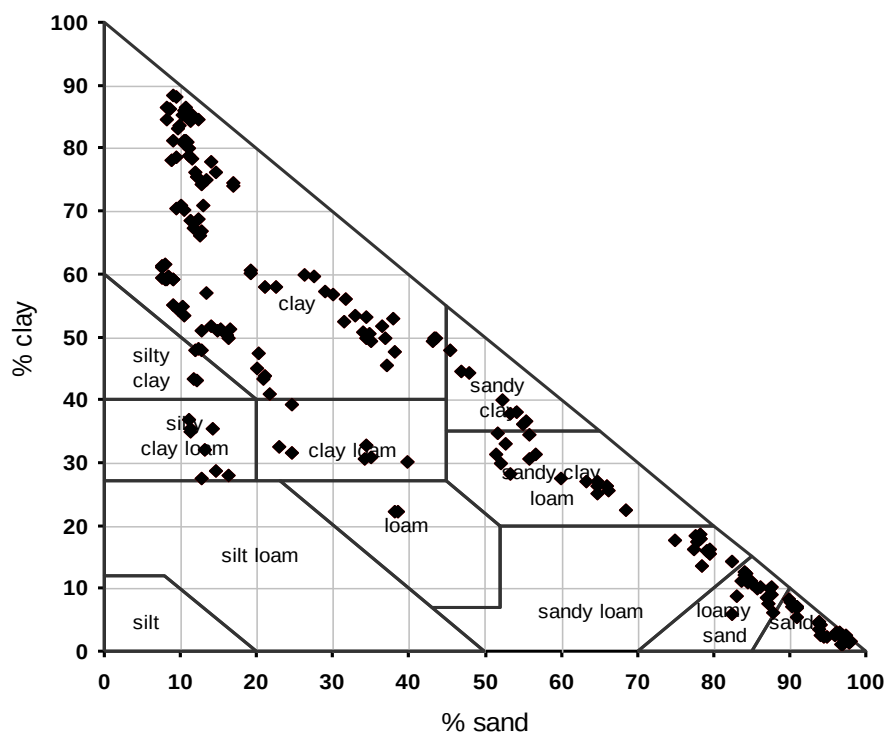


Figure 7-1. Variation of clay (0–2 µm), silt (2–50 µm) and sand (50–2000 µm) of 196 soil samples of the Lower Congo dataset

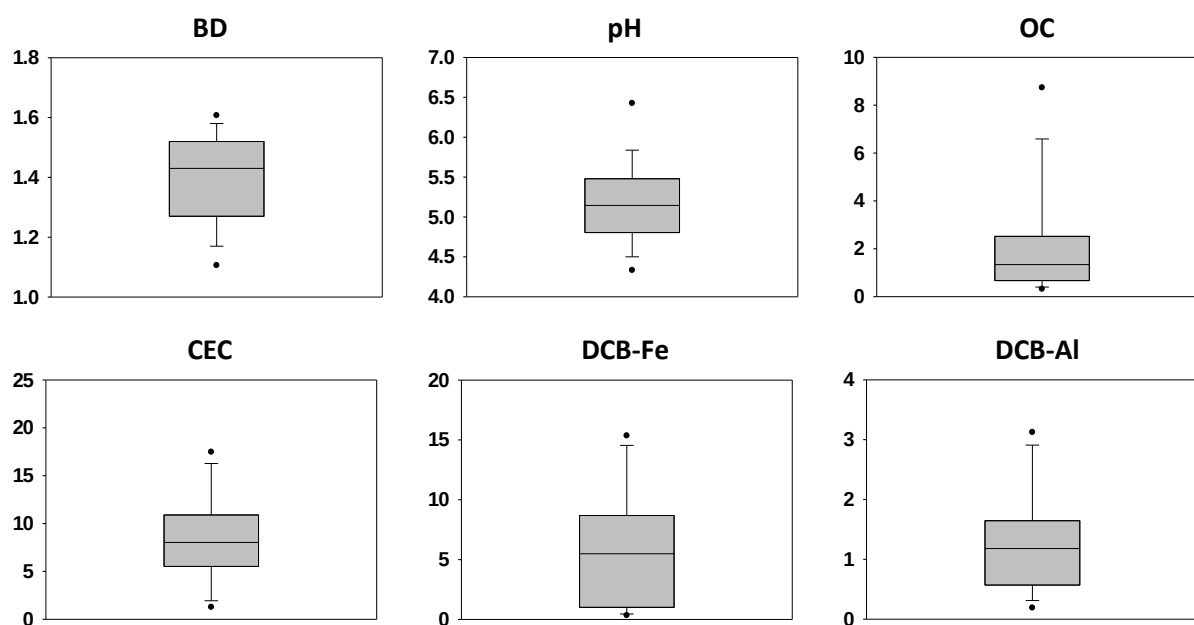


Figure 7-2. Box plots of selected physical and chemical properties of 196 soil samples of Lower Congo. BD is bulk density (Mg m^{-3}), pH is acidity (-), OC is organic carbon content (%), CEC is cation exchange capacity ($\text{cmol}_c \text{ kg}^{-1}$ soil), DCB-Fe is dithionite-citrate-bicarbonate extractable iron (%) and DCB-Al is dithionite-citrate-bicarbonate extractable aluminium (%).

7.2.3 Predictive capability of published bulk density pedotransfer functions on the Lower Congo dataset

In this study, eight multiple linear regression (MLR) PTFs were tested to predict BD of soils of Lower Congo (Table 7–1). The first four selected PTFs were developed using soils of tropical climate regions (Bernoux et al., 1998; Kaur et al., 2002; Benites et al., 2007; Minasny and Hartemink, 2011) and the other BD PTFs were derived from soils of temperate climate regions (Huntington et al., 1989; Manrique and Jones, 1991; Leonavičiūtė, 2000; Hollis et al., 2012). A visual test using 1:1 plots was used to get a quick insight in the predictive capability of BD PTFs derived from ‘tropical’ soils and BD PTFs derived from ‘temperate’ soils.

7.2.4 Development of pedotransfer functions to predict bulk density

Pedotransfer functions for Lower Congo soils were developed based on two different approaches: (1) an equation-based approach where multiple linear regression (MLR) equations are developed using stepwise regression procedure, and (2) an approach based on pattern-recognition called the k-nearest neighbour (k-NN) method.

7.2.4.1 Multiple linear regression pedotransfer functions

In this section, PTFs were developed based on a stepwise regression method and using 196 undisturbed soils from Lower Congo as development (training) dataset. The SAS software (SAS Institute Inc., Cary, NC, USA) was used to develop the MLR PTFs. The selection of the predictors to develop the BD PTFs was done on a hierarchical basis, i.e., using different levels of availability of soil properties that can be considered as potential predictors of BD in the MLR approach (Table 7–2).

7.2.4.2 Pedotransfer functions based on pattern recognition

The pattern-recognition approach used to predict BD for Lower Congo soils is based on the k-NN method. The k-NN technique does not use any predefined mathematical function to estimate a certain response attribute. It does not appear to rely on any stringent assumptions about the underlying data and can adapt to any situation (Hastie et al., 2009). The k-NN approach consists of finding the k number of nearest neighbours from a reference dataset to each soil in the test dataset in terms of their selected input attributes. In ascending order of their similarity distance to the target soil, nearest soils are then sorted. Once the nearest neighbours are identified and sorted, distance-dependent weights are assigned to them and the response attribute (here BD) is formulated and outputted as the weighted average of the response attributes of the selected nearest neighbours. For more theoretical details, the interested reader can refer to Nemes et al. (2006a,b), Chapter 3 and Chapter 6.

Table 7-1. ‘Tropical’ and ‘temperate’ pedotransfer functions used to predict bulk density of Lower Congo soils.†

Source	Output	Geographical domain
‘Tropical’ PTFs		
Bernoux et al. (1998)	$BD = 1.52 - 0.0038 * Cl - 0.05 * OC - 0.045 * pH + 0.001 * Sa$	Brazil (Amazonia)
Kaur et al. (2002)	$\ln(BD) = 0.313 - 0.191 * OC + 0.02102 * Cl - 0.000476 * (Cl)^2 - 0.00432 * Si$	India
Benites et al. (2007)	$BD = 1.5597 - 0.0004 * Cl - 0.0077 * OC - 0.0004 * DCB - Fe$	Brazil
Minasny and Hartemink (2011)	$BD = \frac{100}{\left(\frac{OM}{BD_{om}} + \frac{100 - OM}{BD_{min}} \right)}$ with: $BD_{org} = \text{organic matter bulk density} = 0.224 \text{ Mg m}^{-3}$ $BD_{min} = 0.935 + 0.049 * \ln(\text{depth}) + 0.0055 * Sa + 0.000065 * (Sa - 38.96)^2$	World tropical soils
‘Temperate’ PTFs		
Huntington et al. (1989)	$\ln(BD) = -2.39 - 1.316 * \ln(OM) - 0.167 * [\ln(OM)]^2$	US (New Hampshire)
Manrique and Jones (1991)	$BD = 1.66 - 0.318 * \sqrt{OC}$	US, Hawaii, Puerto Rico, and others
Leonavičiūtė (2000)	$BD = 0.69794 + 0.750636 * EXP(-0.230355 * OC) + 0.0008687 * Sa - 0.0005164 * Cl$	Lithuania
Hollis et al. (2012)	$BD = 0.69794 + 0.750636 * EXP(-0.230355 * OC) + 0.0008687 * Sa - 0.0005164 * Cl$	Europe

† Cl is clay (%), Si is silt (%), Sa is sand (%), BD is bulk density (Mg m^{-3}), OC is organic carbon (%), OM is organic matter (%), CEC is cation exchange capacity ($\text{cmol}_c \text{ kg}^{-1}$ soil), DCB-Fe is dithionite-citrate-bicarbonate extractable iron (%).

Table 7-2. Different levels of input variables that are used as potential predictors in the stepwise regression procedure to develop bulk density pedotransfer functions for soils in Lower Congo.†

Level	Variables used in the stepwise regression process
1	Sand + Silt + Clay (SSC)
2	SSC + pH + OC
3	SSC + pH + OC + CEC
4	SSC + pH + OC + CEC + DCB-Fe + DCB-Al

† SSC is sand, silt and clay, OC is organic carbon, pH is acidity, CEC is cation exchange capacity, DCB-Fe is dithionite-citrate-bicarbonate extractable iron, DCB-Al is dithionite-citrate-bicarbonate extractable aluminium.

The k-NN algorithm used here has been adapted (see Chapter 6) from the variant of the k-NN algorithm developed by Nemes et al. (2006a) to predict water retention of soils from temperate climates. Gharahi Ghehi et al. (2012) applied with success this algorithm to predict BD of soils from Rwanda. In the present study, the k-NN algorithm was adapted and tested for prediction of BD of 196 Lower Congo soils. The Lower Congo dataset was used both as training and test dataset in a leave-one-out (LOO) procedure (Mucherino et al., 2009). The implementation was done in the MATLAB R2010a environment (The MathWorks, Inc., Hill Drive Natick, MA). The selection of the predictors for this k-NN variant was also done on a hierarchical basis. Therefore, 15 sets of input variables were considered in order to investigate the contribution of various predictors to the predictive ability of the k-NN method (Table 7-3).

Table 7-3. Different levels of input variables considered in the k-nearest neighbour algorithm to predict bulk density of soils in Lower Congo.†

Level	Input attributes
1	OC
2	DCB-Fe
3	DCB-Al
4	OC + DCB-Fe + DCB-Al
5	Sand + Silt + Clay
6	SSC + OC
7	SSC + pH
8	SSC + CEC
9	SSC + DCB-Fe + DCB-Al
10	SSC + OC + pH
11	SSC + OC + pH + CEC
12	SSC + OC + DCB-Fe + DCB-Al
13	SSC + CEC + DCB-Fe + DCB-Al
14	SSC + OC + pH + DCB-Fe + DCB-Al
15	SSC + OC + pH + CEC + DCB-Fe + DCB-Al

† SSC is sand, silt and clay, OC is organic carbon, pH is acidity, CEC is cation exchange capacity, DCB-Fe is dithionite-citrate-bicarbonate extractable iron, DCB-Al is dithionite-citrate-bicarbonate extractable aluminium.

7.2.4.3 Statistical criteria to evaluate the predictive ability of the MLR PTFs and the k-NN approach

Four statistical indices were selected to assess the predictive ability of the MLR PTFs and the k-NN algorithm: the mean difference (MD), the mean absolute difference (MAD), the root mean square difference (RMSD) and the coefficient of determination (R^2):

$$MD = \frac{1}{N} \sum_{i=1}^N (\rho_{p_i} - \rho_{m_i}) \quad [7-1]$$

$$MAD = \frac{1}{N} \sum_{i=1}^N |\rho_{p_i} - \rho_{m_i}| \quad [7-2]$$

$$RMSD = \sqrt{\frac{1}{N} \sum_{i=1}^N (\rho_{p_i} - \rho_{m_i})^2} \quad [7-3]$$

$$R^2 = \frac{\left(\sum_{i=1}^N (\rho_{p_i} - \overline{\rho_{p_i}})(\rho_{m_i} - \overline{\rho_{m_i}}) \right)^2}{\sum_{i=1}^N (\rho_{p_i} - \overline{\rho_{p_i}})^2 \sum_{i=1}^N (\rho_{m_i} - \overline{\rho_{m_i}})^2} \quad [7-4]$$

where ρ_{p_i} and ρ_{m_i} are predicted and measured bulk density for soil sample i (Mg m^{-3}), respectively; $\overline{\rho_{p_i}}$ and $\overline{\rho_{m_i}}$ are the means of the predicted and measured bulk density, respectively; and N is the number of samples in the development or training soil dataset.

The overall prediction error in MLR PTFs is generally calculated using the RMSD index (Equation 7–3). In the calculation of RMSD, all members of the development dataset used to derive the equation are considered and therefore $N = 196$. In the k-NN approach, the Lower Congo dataset is used both as the training and the test dataset. Therefore, the RMSD is calculated based on a LOO procedure with $N = 196 - 1 = 195$. As the name suggests, the LOO method is performed by leaving only one sample out of the training dataset: all the samples except the one left out are used as a training dataset, and the model is validated on the sample left out (Mucherino et al., 2009).

7.3 Results and discussion

7.3.1 Predictive ability of published BD PTFs applied on the Lower Congo dataset

Similar to the study conducted in Chapter 5, published ‘temperate’ and ‘tropical’ MLR PTFs have been selected to predict BD of soils in Lower Congo prior to potential development of new PTFs (Table 7–1). The aim of this preliminary evaluation was to show any significant difference in prediction performance between BD PTFs developed based on soils in tropical climates and the ones developed based on soils in temperate climates. All the tested ‘temperate’ as well as ‘tropical’ PTFs clearly underpredicted BD of Lower Congo soils, whereas the ‘temperate’ PTF of Leonavičiūtė (2000) showed a marked overestimation (Figure 7–3).

Published PTFs derived from tropical areas did not perform better than PTFs developed for temperate areas. These results show that the poor prediction performance of existing BD PTFs cannot be only attributed to the geographical location of the soils in the development dataset. Most datasets used to derive the published PTFs were dominated by soils with a high amount of OM which contributed to a lower BD in these soils. The presence of high amounts of OM in the soil favours biological activity

linked to burrowing animals and/or plant roots in various sampling sites that can increase porosity and consequently lead to a decrease of BD in some soil horizons. Another reason of the clear underestimation of BD for Lower Congo soils by the tested PTFs could be that some of these PTFs were developed based exclusively on forest soils with a limited number of samples from deep horizons whereas in this study, a large proportion of non-forest soils and subsoil samples are included. Non-forest soils such as agricultural soils are subjected to human activities such as tillage that can lead to soil compaction. The latter results in an increase of soil BD which is more pronounced in the surface horizons. Jalabert et al. (2010) found that soil fauna, roots, compaction due to management practices can cause local variations in BD, particularly in the surface horizons.

In developing models to predict BD of Lower Congo soils, we followed the recommendation of De Vos et al. (2005) relating to the inclusion of a large proportion of subsoil samples when formulating BD PTFs. This is particularly important for soils from the humid tropics which can be very deep, sometimes extending beyond 10 m depth (Schwartz and Namri, 2002). On the other hand, the PTF of Leonavičiūtė (2000) considered here was specifically developed to predict BD of consolidated B horizons of soils of Lithuania. These horizons were designated as B_w , B_g , B_t , i.e., mineral structured subsurface horizons either being freely drained, gleyed, or illuvial. Soil samples collected from these horizons and used to derive the PTF were characterized by relatively high BD. This can explain the clear overestimation of BD when this PTF is applied to Lower Congo soils.

The results of this preliminary evaluation study show that BD is a soil property that is difficult to predict with a satisfactory level of reliability with a PTF that was developed based on soils different from the soils of interest. This observation is in agreement with Nemes et al. (2010) who reported that the safest option is to use a PTF that has been developed using data of the intended application region. In addition, Suuster et al. (2011) indicated that some PTFs work only for a very restricted area and under certain conditions (or even within one geographic location).

Empirical equations like those based on MLR are mostly applicable to soils located in sites similar to those used in deriving the BD PTFs and therefore cannot be easily extrapolated.

As a way to further improve the prediction of PTFs, authors such as Pachepsky and Rawls (1999), Leonavičiūtė (2000), Nemes et al. (2010), Hollis et al. (2012), recommended a preliminary grouping of soils prior to the development of PTFs. However, grouping of soils based on various criteria can only be really beneficial when a large dataset of various soil samples is considered. Hollis et al. (2012) collected 1,545 soil samples from England and Wales in order to form statistically significant groups related to various characteristics such as parent material, depth, and OC content.

The results found in this preliminary evaluation of published PTFs and the lack of BD data for soils of Central Africa (e.g., the D.R. Congo and the R. Congo) for, e.g., carbon stock assessments, clearly show the need of developing reliable BD PTFs. In the same vein, as suggested by several authors (e.g., Kaur et al., 2002; De Vos et al., 2005; Nemes et al., 2010), PTFs should be developed for this particular region in order to reduce uncertainties in predicting BD and therefore in accurately assessing carbon stocks in Central Africa.

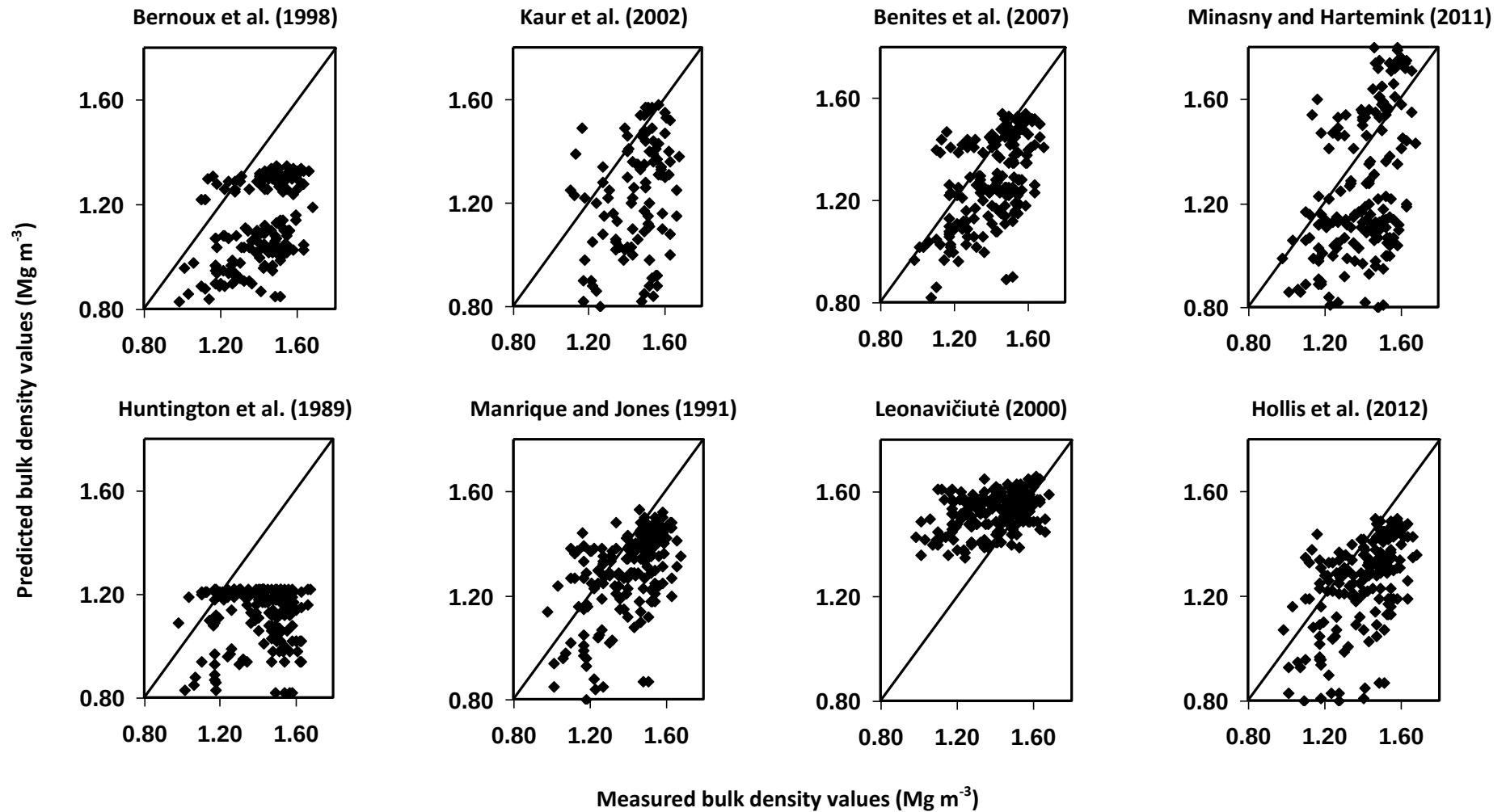


Figure 7–3. Measured versus predicted bulk density values of 196 soil samples from Lower Congo using four ‘tropical’ and four ‘temperate’ pedotransfer functions.

7.3.2 Hierarchical multiple linear regression PTFs to predict bulk density

With limited to more extended sets of potential predictors considered in a regression procedure (see Table 7–2), two MLR PTFs were selected to predict BD of soils from Lower Congo. On the one hand, out of six available soil properties (clay, silt, sand, pH, OC and CEC), only three predictors were retained: clay, sand and OC. On the other hand, out of eight soil properties made available for the regression analysis (clay, silt, sand, OC, pH, CEC, DCB-Fe and DCB-Al), only three predictors were found to be significant: DCB-Fe, DCB-Al and OC. Table 7–4 shows the correlation matrix between the properties of the soil considered in this study. The BD PTFs that were derived by the regression method are presented in Table 7–5. Plots of residuals versus predictors for PTF-1 (clay, sand and OC) and PTF-2 (DCB-Fe, DCB-Al and OC) did not exhibit any systematic structure (Figure 7–4). This indicates that the model fits the data well. Using logarithmic or squared terms for the various potential predictors did not bring any substantial improvement in the prediction quality. Performing a subdivision of data into groups by soil depth, i.e., topsoil data and subsoil data did not improve the prediction of BD for this dataset. Therefore, soil depth was not considered as a potential predictor of BD. Nemes et al. (2010) found that grouping samples by horizon notation did not yield significant improvement in the prediction quality due to the great variability of BD within horizons of the same notation. Suuster et al. (2011) noted that for most PTFs found in literature, sampling depth was not considered as a possible variable for predicting soil BD.

Table 7-4. Pearson's correlation matrix between the various potential predictors and measured soil bulk density (N=196).

	Clay	Silt	Sand	pH	OC	CEC	DCB-Fe	DCB-Al	BD
Clay	1.00								
Silt	0.210**	1.00							
Sand	-0.900**	-0.615**	1.00						
pH	-0.072	-0.237**	0.164*	1.00					
OC	0.461**	0.158*	-0.442**	-0.040	1.00				
CEC	0.548**	0.465**	-0.649**	-0.091	0.534**	1.00			
DCB-Fe	0.902**	0.083	-0.764**	-0.034	0.544**	0.532**	1.00		
DCB-Al	0.846**	0.032	-0.697**	0.004	0.348**	0.534**	0.870**	1.00	
BD	-0.421**	0.037	0.323**	-0.107	-0.382**	-0.417**	-0.426**	-0.530**	1.00

* shows significant correlation at 0.05 significance level

** shows significant correlation at 0.01 significance level

Table 7-5. Regression relationships between physico-chemical properties and bulk density for 196 soil samples from Lower Congo.†

Model	Intercept	Clay	Sand	OC	DCB-Fe	DCB-Al	MAD	RMSD	R ²
		%	%	%	%	%	(Mg m ⁻³)	(Mg m ⁻³)	(-)
PTF-1	1.64581	-0.00362	-0.0016	-0.01580			0.110	0.137	0.244
PTF-2	1.54373			-0.02287	0.01518	-0.15198	0.100	0.126	0.369

† PTF is pedotransfer function, OC is organic carbon, DCB-Fe is dithionite-citrate-bicarbonate extractable iron, DCB-Al is dithionite-citrate-bicarbonate extractable aluminium, MAD is mean absolute difference RMSD is root mean square difference; R² is the coefficient of determination (-).

The two PTFs were compared based on the aforementioned statistical indices in order to identify the best equation as well as the best predictors of BD for soils of Lower Congo. From Table 7–5, it can be seen that both PTF-1 and PTF-2 presented RMSD values less than 0.140 Mg m^{-3} with the highest MAD (0.110 Mg m^{-3}) and RMSD (0.137 Mg m^{-3}) for PTF-1 with clay, sand and OC as explanatory variables, as compared to MAD (0.100 Mg m^{-3}) and RMSD (0.126 Mg m^{-3}) for PTF-2 with OC, DCB-Fe and DCB-Al as predictor variables. These RMSD values are comparable and often lower than RMSD values found by other authors when developing BD PTFs for their ‘tropical’ soils (e.g., Benites et al., 2007: 0.190 Mg m^{-3} ; Minasny and Hartemink, 2011: 0.200 Mg m^{-3}).

On the other hand, R^2 values were found to be relatively low (0.244 for MLR PTF-1 and 0.369 for MLR PTF-2). In previous works on prediction of soil BD (e.g., Bernoux et al., 1998; Benites et al., 2007; Minasny and Hartemink, 2011), low values of R^2 (or R^2 -adj) between 0.30 and 0.46 were also found for some derived BD PTFs.

As can be seen in Figure 7–5, no strong linear relationship was found between BD and other physico-chemical properties of soils of the Lower Congo dataset. However, one can still observe an inverse relationship between BD and clay, OC, DCB-Fe and DCB-Al and a direct relationship between BD and sand content in the Lower Congo dataset. Particle density of sand is generally higher than that of clay and sandy soils are known to lack structure and to have higher BD than clayey soils (Hillel, 1998). On the other hand, as binding agents of soil particles, OC, DCB-Fe and DCB-Al favour aggregation and therefore contribute to a higher porosity and consequently a decrease in soil BD.

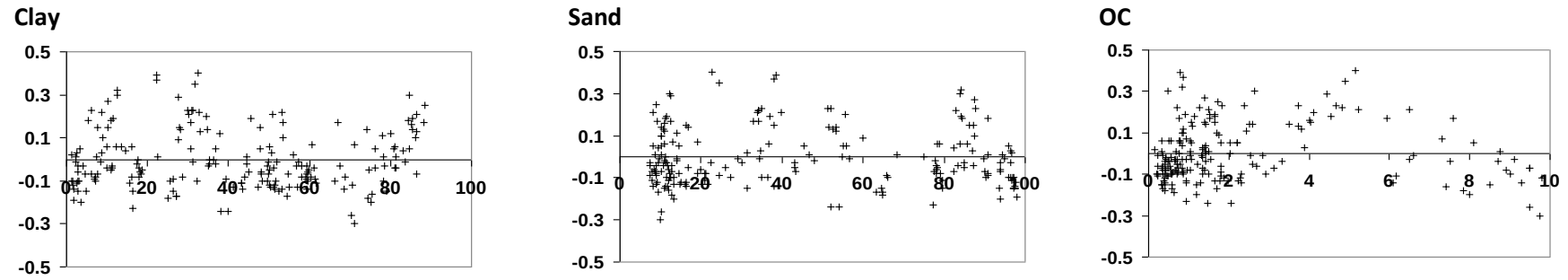
The lower RMSD values yielded by MLR PTF-2 as compared to PTF-1 indicate that OC, DCB-Fe and DCB-Al are important predictors of soil BD. Moreover, it was found that DCB-Fe and DCB-Al can constitute better explanatory variables of BD than texture. This can be explained by the fact that the Fe oxides are the main responsible drivers of the micro-aggregation of soil particles in highly weathered soils, dominated by kaolinite.

Bronick and Lal (2005) indicated that the rate and stability of aggregation generally increases with OC, clay surface area and CEC. Crystalline and especially noncrystalline Al and Fe oxides play an important role in the stabilization of soil aggregates (Oades and Waters, 1991; Barral et al., 1998). In the humid tropics, the tendency for Al to replace Fe in Fe oxide minerals diminishes the crystallinity of the oxides and the size of the particles, leading to a significant increase in surface area and the development of variable charges (Van Ranst et al., 1998; Baert et al., 1999). Six et al. (2000) indicated that the interaction of Al and Fe with kaolinite can synergistically promote aggregation with limited impact on soil OC. On the other hand, Molina et al. (2001) observed that the synergistic interaction between Al oxides and hydroxides with soil OC and dispersible clay, improved stability of soil aggregates. Al hydroxide is typically present in soils as the mineral gibbsite $[\text{Al}(\text{OH})_3]$, whereas goethite (FeOOH) and hematite (Fe_2O_3) are the most common Fe oxides found in highly weathered soils in Central Africa. Their small particle size and positive charge make them effective as binding agents on negatively charged clay minerals. Denef et al. (2002) found that soils dominated by variable charge clay minerals such as 1:1 clay (e.g., kaolinite) and oxides have higher aggregation at lower OC levels. This is particularly true for the Lower Congo dataset containing samples of highly weathered soils mostly collected in deeper horizons with limited OC content.

As the DCB-Fe and DCB-Al contents of strongly weathered soils have an influence on their structure, they are closely associated with the resulting BD of these soils. The results of this study show that when information on DCB-Fe and/or DCB-Al data are available in soil survey reports (as it is the case for Lower Congo), these soil characteristics should be used as predictors of BD because they are better related to the structure of strongly weathered soils. The results of this study show that PTFs based on

OC, DCB-Fe and DCB-Al have a good potential to predict BD of highly weathered soils when data obtained by conventional methods are missing. This confirms the findings of Benites et al. (2007) for soils from Brazil, who also used DCB-Fe as a BD predictor. However, more studies should be performed in other regions or countries of the humid tropics to confirm these findings.

(a)



(b)

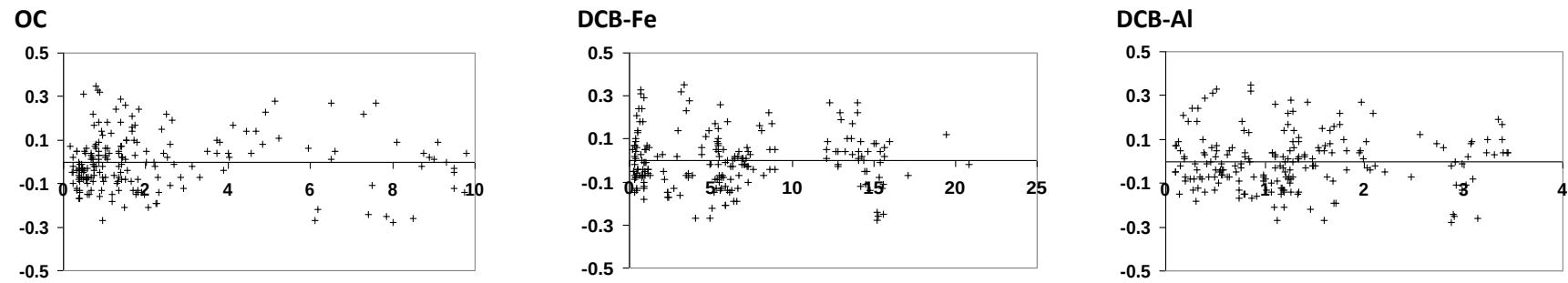


Figure 7-4. Plots of residuals versus predictors for (a) PTF-1 (BD vs. clay, sand and OC) and (b) PTF-2 (BD vs. DCB-Fe, DCB-Al and OC). Sand and clay contents are in %, BD is bulk density (Mg m^{-3}), OC is organic carbon content (%), DCB-Fe is dithionite-citrate-bicarbonate extractable iron (%) and DCB-Al is dithionite-citrate-bicarbonate extractable aluminium (%).

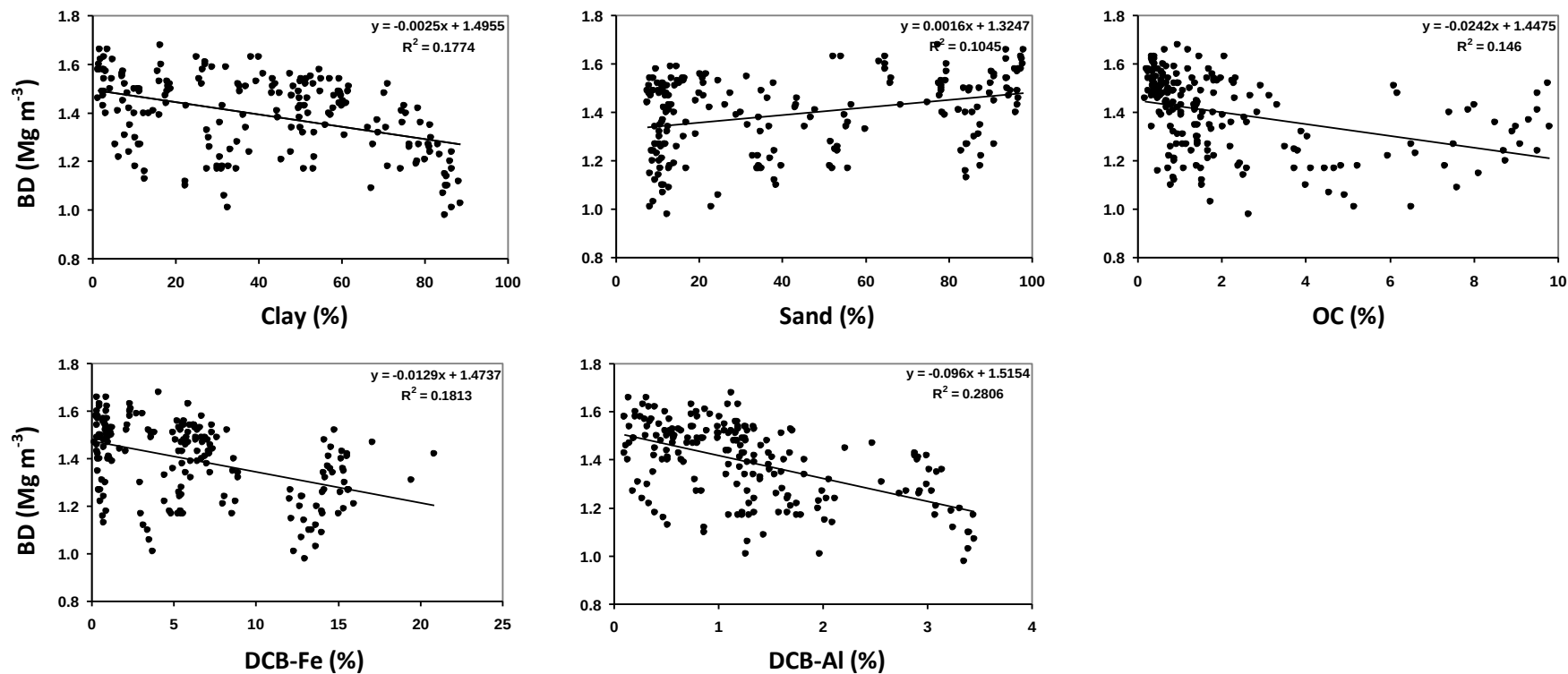


Figure 7–5. Observed bulk density versus clay, sand, OC, DCB-Fe and DCB-Al contents (%) of 196 soil samples of the Lower Congo dataset.

7.3.3 Pedotransfer functions based on pattern-recognition to predict bulk density

A second approach, based on pattern-recognition, was tested to predict BD of soils in Lower Congo. This non-parametric approach using a k-NN variant was applied by Gharahi Ghehi et al. (2012) to predict BD of tropical rainforest topsoils in Rwanda. One of the main advantages of using this approach compared to the MLR approach is that no equation needs to be redeveloped when new data are added to the dataset (Nemes et al., 2006a,b). As previously mentioned, the Lower Congo dataset is used here both as training and test dataset in a LOO procedure. For the k-NN approach, first 27 different sets of input BD predictors were tested. It was found that in the present application, DCB-Fe and DCB-Al can be used interchangeably or both included as predictors without sensitive loss of accuracy. Therefore, for sake of parsimony, only 15 sets of input variables (i.e., input levels) are presented along with a statistical summary of the prediction performance of the k-NN method for each input level (Table 7–6). Among the 15 combinations of inputs considered, ten input combinations yielded RMSD values less than 0.100 Mg m⁻³. The highest RMSD values of 0.134–0.139 Mg m⁻³ were found for levels 1, 2 and 3 with OC, DCB-Fe and DCB-Al as single predictor, respectively, and the lowest RMSD value of 0.093 Mg m⁻³ was found for levels 12 and 13 with SSC, OC or CEC, DCB-Fe and DCB-Al as input variables in the k-NN algorithm.

Table 7-6. Prediction performance in terms of MD, RMSD and R², for the k-nearest neighbour algorithm with optimized settings and using 15 levels of input attributes.†

Level††	Input attributes	MD	RMSD	R ²
1	OC	-0.006	0.139	0.263
2	DCB-Fe	-0.002	0.134	0.302
3	DCB-Al	0.001	0.139	0.252
4	OC + DCB-Fe + DCB-Al	0.002	0.108	0.537
5	Sand + Silt + Clay	0.005	0.102	0.583
6	SSC + OC	0.009	0.096	0.634
7	SSC + pH	0.001	0.100	0.601
8	SSC + CEC	0.006	0.098	0.617
9	SSC + DCB-Fe + DCB-Al	0.004	0.098	0.613
10	SSC + OC + pH	0.005	0.096	0.635
11	SSC + OC + pH + CEC	0.000	0.096	0.636
12	SSC + OC + DCB-Fe + DCB-Al	0.005	0.093	0.660
13	SSC + CEC + DCB-Fe + DCB-Al	0.004	0.093	0.656
14	SSC + OC + pH + DCB-Fe + DCB-Al	0.003	0.095	0.639
15	SSC + OC + pH + CEC + DCB-Fe + DCB-Al	-0.002	0.095	0.640

† OC is organic carbon (%), DCB-Fe is dithionite-citrate-bicarbonate extractable iron (%), DCB-Al is dithionite-citrate-bicarbonate extractable aluminium (%), MAD is mean absolute difference (Mg m⁻³); RMSD is root mean square difference (Mg m⁻³); R² is the coefficient of determination (–).

†† Levels 1 to 15 represent selected levels of availability of soil information in the development (training) dataset.

These results indicate that using only one predictor, such as OC, to predict BD for Lower Congo soils can lead to sensitive bias in the estimation. This is somewhat different from the results obtained by Gharahi Ghehi et al. (2012) for ‘tropical’ soils in Rwanda. The authors found that using OC as sole predictor in their k-NN algorithm was appropriate for estimating topsoil BD in tropical forests. The contrasting results between both studies can be partly explained by the difference in land cover (vegetation) types of the collection sites on the one hand and the depth at which most of the soil samples were collected on the other hand. The training dataset used by Gharahi Ghehi et al. (2012) essentially

consists of forest topsoil samples taken in Rwanda whereas most soil samples of the Lower Congo dataset were collected under savannah with a dominance of subsoil samples and presence of coarse textured soils with limited OC content. In previous studies, several authors (e.g., Adams, 1973; Gosselink et al., 1984; Ruehlmann and Körschens, 2009) found OC to be the main controlling factor of BD. However, Jalabert et al. (2010) indicated that the relative influence of soil OC on BD is often greater in forest soils because of their high OC content.

Considering the results obtained with the k-NN method when using only OC as predictor of BD for Lower Congo soils ($\text{RMSD} = 0.139 \text{ Mg m}^{-3}$), a sensitive improvement of the predictive quality of the k-NN method has been obtained by adding other predictors such as texture and DCB-Fe and/or DCB-Al ($0.093 \text{ Mg m}^{-3} < \text{RMSD} < 0.108 \text{ Mg m}^{-3}$). Using the pattern-recognition approach, the importance of OC, DCB-Fe and DCB-Al as predictors of BD of strongly weathered soils was put in evidence.

A comparison between measured and predicted values of BD by the k-NN approach using different sets of input variables is illustrated in Figure 7–6. It can be seen that the prediction performance of the k-NN technique is quite good compared to the MLR technique. When considering similar inputs as MLR PTF-1 (clay, sand and OC) and MLR PTF-2 (OC, DCB-Fe and DCB-Al), the overall prediction performance of level 6 (clay, silt, sand and OC) and level 4 (OC, DCB-Fe and DCB-Al) of the k-NN PTFs is sensitively better than MLR PTF-1 and MLR PTF-2, respectively (Figure 7–7). The high predictive ability of the k-NN approach can be explained by the fact that soil samples in the Lower Congo dataset which are the closest to the target soil will get the highest weights in the calculation of the final output by the algorithm as explained in Chapter 6. Therefore, only the most relevant soils will be actually taken into consideration in the prediction of BD. These results demonstrate that the k-NN technique based on pattern-recognition is an interesting alternative for predicting BD of highly weathered soils in Central Africa.

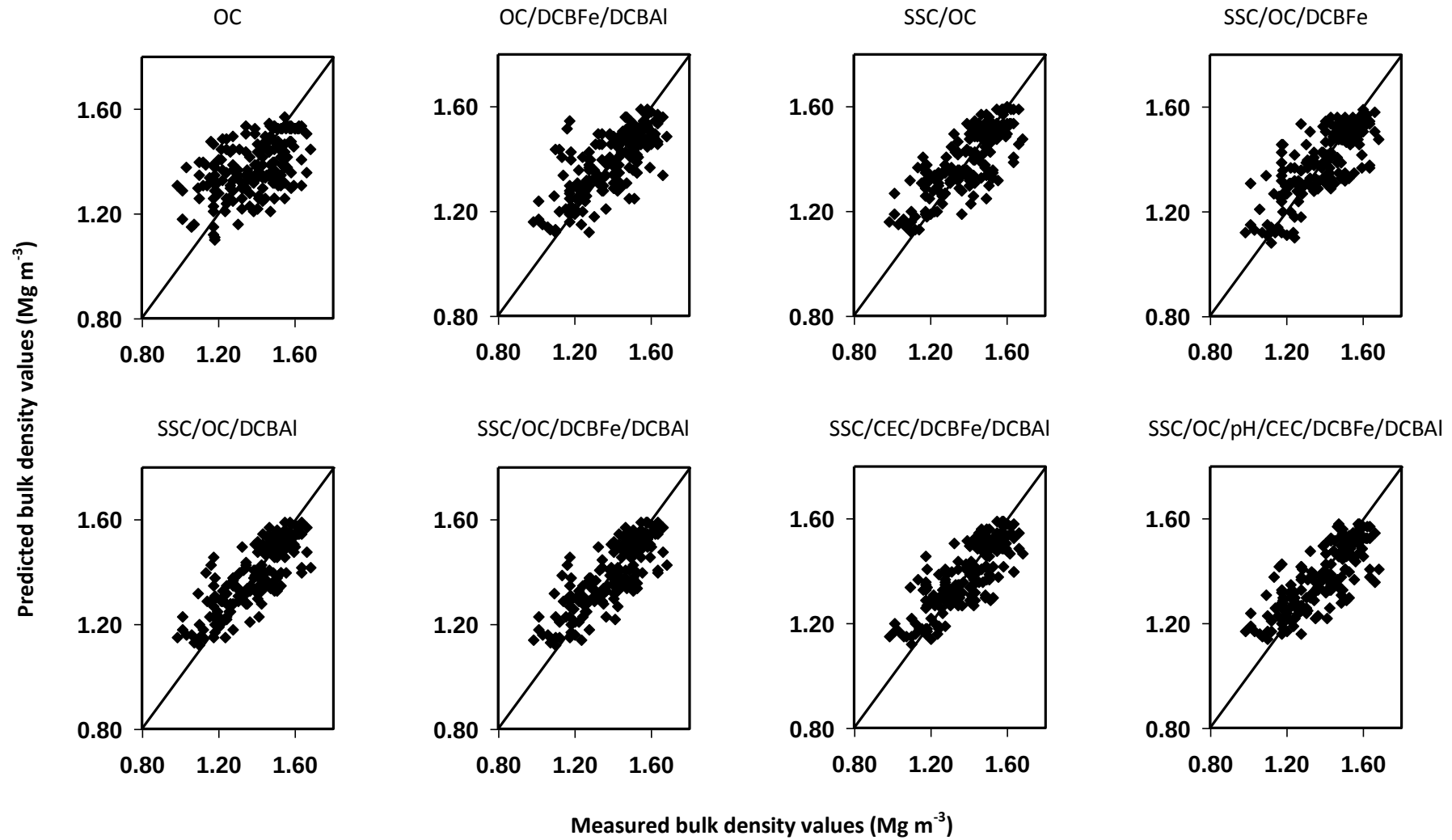


Figure 7–6. Measured versus predicted bulk density values of 196 soil samples from Lower Congo by the k-nearest neighbour method with eight different set of input attributes.

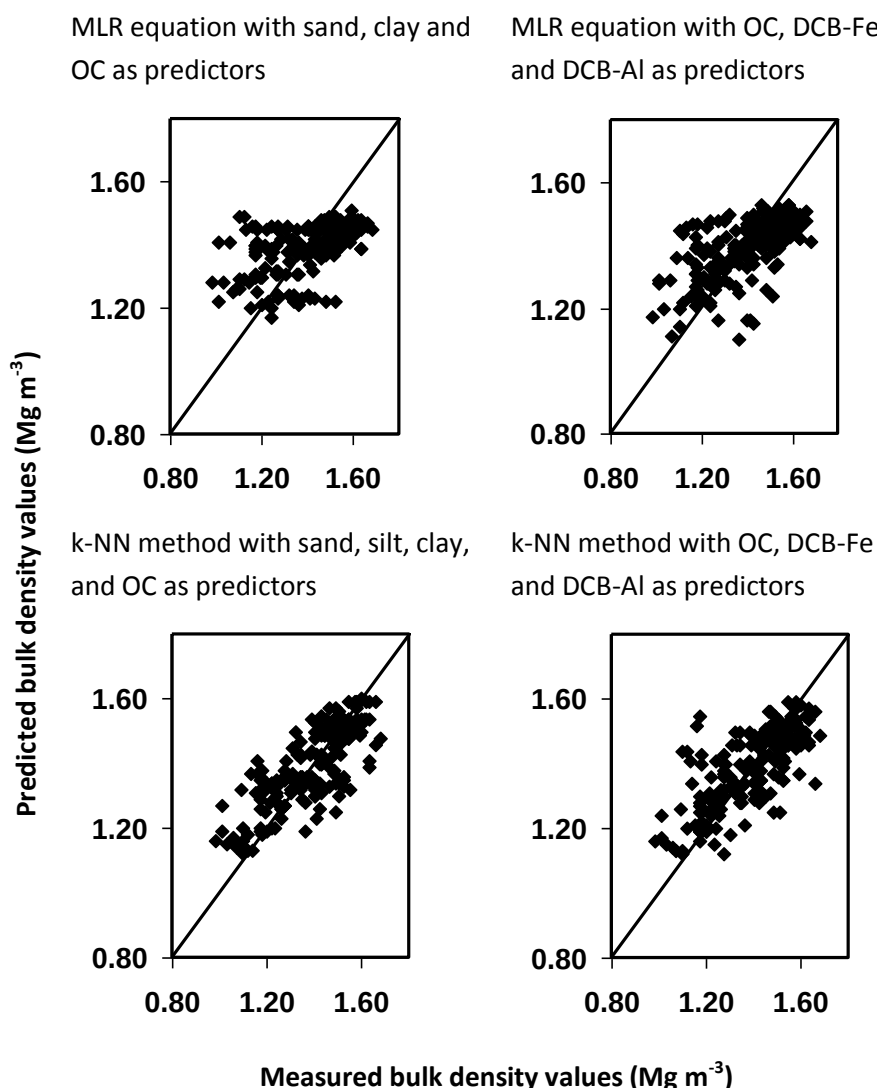


Figure 7–7. Measured versus predicted bulk density values of 196 soil samples from Lower Congo by the multiple linear regression (MLR) and the k-nearest neighbour (k-NN) methods.

7.4 Conclusion

Data on BD are often missing in most soil databases of Central Africa. To respond to this paucity of legacy data, PTFs are generally used. In the present study, the predictive abilities of four tropical and four temperate published BD PTFs were tested on a collection of highly weathered soil samples from Lower Congo. Great discrepancies were found between the observed BD values and the predicted values making published PTFs impractical for soils of Lower Congo. Therefore, new PTFs have to be developed to predict BD of strongly weathered soils of Lower Congo.

In order to respond to this need, two main approaches were used in this study: the equation-based approach and the pattern-recognition approach. In the first approach, different PTFs in the form of MLR equations were formulated to predict BD based on four different levels of availability of soil information. In the second approach, a k-NN algorithm was used with 15 different sets of input attributes to provide estimates of BD for Lower Congo soils in a LOO procedure. The results in terms of RMSD values of both approaches were better than those of other published PTFs developed based on temperate and tropical soil datasets.

The relative importance of DCB-Fe and DCB-Al in the improvement of BD prediction in both approaches was demonstrated along with other commonly used soil characteristics such as OC and

texture. Additionally, the predictive abilities of the MLR PTFs and the k-NN PTFs were compared. With the same set of input predictors, the k-NN method performed the best with RMSD values of 0.096 Mg m^{-3} when using clay, sand and OC content, and 0.108 Mg m^{-3} with OC content, and DCB-Fe and DCB-Al, compared to the MLR method with RMSD values of 0.137 and 0.126 Mg m^{-3} , respectively. As indicated by Nemes et al. (2006a,b) and by the results of the previous study (see Chapter 6), the k-NN method can therefore be considered as a competitive alternative to more classical MLR PTFs due its better prediction capability and its flexibility to incorporate new data without the burden to redevelop new equations. This is particularly advantageous in developing countries where soil databases for assessment of carbon stocks are at present sparse, though slowly developing.

Based on their needs and their familiarity towards one of the two approaches, potential users interested in predicting BD of strongly weathered soils of Lower Congo in particular and of Central Africa in general can immediately use the proposed equations or later apply the k-NN algorithm which will be incorporated in a flexible user-friendly software using, e.g., the present Lower Congo dataset as a training dataset. In future studies related to PTFs, other soil types found in Central Africa such as swelling-shrinking soils (Vertisols) and volcanic soils (Andosols) should also be taken into account with their specific properties. Moreover, other soil information like categorical soil structural data based on the profile descriptions should be investigated as potential predictors of BD of soils of the humid tropics.

8 Hierarchical pedotransfer functions to predict water retention of soils in Lower Congo

8.1 Introduction

Hydraulic properties of soils are important parameters used in research and applications related to pedological, environmental and agricultural studies as well as other soil-related disciplines. They are used as inputs in various models dealing with climate change, estimation of crop yield, transport of solutes and pollutants among others. Soil water retention is a major soil hydraulic property that governs soil functioning in ecosystems and greatly affects soil management (Cornelis et al., 2001; Rawls et al., 2003). Soil can bind water by means of the interactive capillary and adsorptive forces. Both forces result in what is called matric potential or matric suction Ψ (Hillel, 1998). The soil water retention curve (SWRC), also called soil water characteristic curve or soil water release curve, describes the functional relationship between the matric potential of the soil Ψ and its gravimetric or volumetric water content W (Hillel, 1998; Jury and Horton, 2004).

In practice, the measurements of hydraulic properties such as the SWRC are time-consuming, laborious and expensive, especially for environmental studies at a regional scale (Vereecken, 1995; Pachepsky et al., 2006; Guber et al., 2006). Therefore, alternative methods, called pedotransfer functions (PTFs), have been developed to estimate these hydraulic properties based on easy-to-measure soil properties such as soil texture (sand, silt and clay), bulk density (BD), organic matter/organic carbon (OM/OC) content and/or other data routinely measured or registered in soil surveys (Bouma and van Lanen, 1987; Bouma, 1989; van den Berg et al., 1997). However, Tomasella et al. (2000) wrote that most published PTFs have been developed using data from soils of temperate regions, only very few ones are dedicated to PTFs for soils in the humid tropics.

We observed that most of the PTFs developed for soils in the (sub)humid tropics were point-based PTFs (see Chapter 3). Many authors (Pachepsky et al., 1996; Tomasella et al., 2003; Dashtaki et al., 2010; Vereecken et al., 2010) noted that the point-based PTFs predicted better water retention than the parameter-based PTFs. This can be explained by the fact that water retention is controlled by different soil properties depending on the level of soil matric potentials. The point-based PTFs allow for more appropriate independent variables to describe the water content variation than do the parameter-based PTFs.

In Chapter 4, ten different particle size distribution (PSD) models were tested to achieve compatibility between different soil textural classification systems applied in various tropical countries. These PSD models showed to be useful when a given PTF needs to be applied on a certain soil dataset in which the texture classification system used is different from the PTF development dataset.

Based on a limited dataset, we performed a preliminary evaluation of published PTFs to predict water retention at -33 kPa and -1500 kPa of soils in Lower Congo (see Chapter 5). The results of this study showed clearly that there is a need to develop water retention PTFs for this region based on a larger dataset of undisturbed samples.

In Chapter 6, we demonstrated the potential of the non-parametric k-nearest neighbour (k-NN) approach to predict water retention of soils of the humid tropics using an international dataset of 'tropical' soils as training dataset and selected soils from Lower Congo as test dataset. In our study, we used several soil properties as inputs: sand, silt, clay, BD, OC, pH and cation exchange capacity (CEC). The k-NN method based on pattern-recognition, proved to give comparable results to the classical PTFs based on multiple linear regression (MLR) equations, and offered additional advantages such as flexibility and estimation of uncertainty.

We indicated (see Chapter 3) that dithionite-citrate-bicarbonate extractable iron (DCB-Fe) and aluminium (DCB-Al) have been used with success by van den Berg (1996) and van den Berg et al.

(1997) in predicting water retention of strongly weathered soils. In our preliminary study (see Chapter 5), DCB-Fe was found to be a potential predictor for water retention of highly weathered soils in Lower Congo. However, it is quite surprising to see that no additional investigations were conducted to confirm the potential of these soil properties in the improvement of the PTF prediction of water retention of soils under the humid tropics. One of the reasons may be that these properties are not easily available in some soil survey databases. However, it should be noted that the essential function of a PTF is to ‘translate data we have into what we need’ (Bouma, 1989). In that vein, Tomasella et al. (2000) successfully used moisture equivalent as a potential predictor of water retention of soils from Brazil. Moisture equivalent is defined as the water content remaining in a sample (fraction < 2 mm) after centrifuging at 2,400 rpm for 30 min, generally expressed in gravimetric units. This is a soil property commonly determined by the Brazilian soil surveyors and is therefore available as soil information. Similarly, information on Fe and Al content of soils extractable by dithionite-citrate-bicarbonate (DCB-Fe and DCB-Al) are largely available in various soil survey reports for Lower Congo. Availability of these soil properties combined with the results found previously by van den Berg (1996), van den Berg et al. (1997) and this study (see Chapter 5) justify the effort of investigating their relative importance as potential predictors of water retention of soils of Lower Congo. Using a larger dataset of soils from Lower Congo, both as test and training dataset in the present work, the potential of DCB-Fe and DCB-Al as predictors of soil water retention at various matric potentials can also be investigated with a non-parametric approach.

The objectives of this study are: (1) to develop hierarchical PTFs, i.e., PTFs with limited to more extended sets of predictors, based on a large dataset of undisturbed soil samples to predict water retention of soils of Lower Congo; (2) to use a parametric (equation-based) and a non-parametric (pattern-recognition) approach in the process of deriving the water retention PTFs for soils of Lower Congo; (3) to investigate the possibility of using DCB-Fe and/or DCB-Al as additional variables to enhance the prediction capability of the derived PTFs based on the two aforementioned approaches; and (4) to compare prediction performance of the locally developed MLR PTFs with other published MLR PTFs.

8.2 Materials and Methods

8.2.1 Study area: Lower Congo

The study area has been described in Chapter 2.

8.2.2 The Lower Congo soil dataset

A dataset based on 196 undisturbed soil samples was used both as training and test (validation) dataset to develop water retention PTFs for Lower Congo soils. This dataset is the same used in Chapter 7 to derive BD PTFs for Lower Congo soils. These soils are mainly highly weathered soils of the humid tropics classified as Ferralsols, Acrisols, Nitisols and Arenosols (IUSS Working Group WRB, 2006). Soil samples were collected under different land uses (savannah, forest, agricultural fields and in old quarries). The SWRC was determined from the wet to the dry range at eight different matric potentials: -1, -3, -6, -10, -20, -33, -100 and -1500 kPa. The hanging water-column method was used for matric potentials between -1 and -10 kPa using the sand box apparatus (Eijkkelkamp Agrisearch Equipment, Giesbeek, the Netherlands), whereas for matric potentials between -20 and -1500 kPa, pressure chambers (Soil Moisture Equipment, Santa Barbara, CA) were used, following the procedures described in Cornelis et al. (2005). The physico-chemical characteristics of all soil samples (fine earth) were determined using standard methods described in Van Ranst et al. (1999) on the soil samples previously used for SWRC measurements. Selected physico-chemical soil properties

were: sand (50–2000 μm), silt (2–50 μm), and clay content ($< 2 \mu\text{m}$) according to the USDA classification system (Soil Survey Staff, 1951) and determined with the pipette method of Köhn (1929), BD, OC, pH, CEC, DCB-Fe and DCB-Al.

8.2.3 Development of PTFs to predict water retention

The PTFs were developed based on two different approaches: (1) an equation-based approach where multiple linear regression (MLR) equations are developed using a stepwise regression procedure, and (2) an approach based on pattern-recognition called the k-nearest neighbour (k-NN) method.

8.2.3.1 Multiple linear regression PTFs

In the present study, eight point-based PTFs, i.e., functions that predict the water content at selected matric potentials, were developed based on a stepwise regression method (with probability $p < 0.05$ for entry of the predictor variable) and using 196 soils from Lower Congo as development (training) dataset. Non-significant predictors were left out in the forward stepwise process. The SAS 9.2 software (SAS Institute Inc., Cary, NC, USA) was used to develop the MLR PTFs. The selection of the predictors submitted to the regression process was done following a hierarchical approach, i.e., considering different levels of availability of soil properties that can be used as potential predictors in the MLR approach as shown in Table 8–1. Schaap and Leij (1998) and Schaap et al. (2001) developed their artificial neural networks PTFs based on a similar hierarchical approach with limited to more extended sets of predictors.

Table 8-1. Different levels of availability of soil properties used as potential predictors in the stepwise regression procedure to develop water retention pedotransfer functions for soils in Lower Congo.†

Level	Variables used in the stepwise regression process
1	Sand + Silt + Clay (SSC)
2	SSC + pH + OC
3	SSC + pH + OC + CEC
4	SSC + pH + OC + CEC + DCB-Fe + DCB-Al
5	SSC + BD
6	SSC + BD + pH + OC + CEC
7	SSC + BD + pH + OC + CEC + DCB-Fe + DCB-Al

† SSC is sand, silt and clay (texture), BD is bulk density, OC is organic carbon, pH is acidity, CEC is cation exchange capacity, DCB-Fe is dithionite-citrate-bicarbonate extractable iron, DCB-Al is dithionite-citrate-bicarbonate extractable aluminium.

8.2.3.2 Pedotransfer functions based on pattern-recognition

The pattern-recognition approach used to predict water retention for Lower Congo soils is based on the k-nearest neighbour (k-NN) method. The k-NN algorithm used in this study to predict water retention has been adapted from the variant developed by Nemes et al. (2006a) using a database of soils from temperate climates. In Chapter 6, we used the same algorithm but expanded it to use more input attributes than only texture, BD and OM, and more output attributes than only water content at -33 kPa and -1500 kPa . The design parameters of the algorithm had been re-evaluated (see Chapter 6) based on a training dataset of soils from the tropics (a subset of the IGBP-DIS dataset collected by ISRIC World Soil Information, the Netherlands) and using 139 selected Lower Congo soils as test dataset in order to better represent soils of humid tropical climates. These 139 soils are part of the current dataset used in this study.

The k-NN technique does not use any predefined mathematical function to estimate a certain response attribute, like classic MLR PTFs do, and does not appear to rely on any stringent assumptions about the underlying data (Hastie et al., 2009). The k-NN approach consists of finding the k number of nearest neighbours from a reference dataset to each soil in the test dataset in terms of their selected input attributes. In ascending order of their similarity distance to the target soil, nearest soils will then be sorted. Once the nearest neighbours are identified and sorted, distance-dependent weights are assigned to them and the response attribute is formulated and outputted as the weighted average of the response attributes of the selected nearest neighbours. Therefore, the predicted water retention at a given matric potential corresponds to the (distance-dependent) weighted sum of observed water retention values of the selected nearest neighbours.

For more theoretical details about the k-NN variant used in this study, the interested reader can refer to Nemes et al. (2006a,b) and to Chapter 6. The implementation of the k-NN algorithm was done in the MATLAB R2010a environment (The MathWorks, Inc., Hill Drive Natick, MA, USA). The selection of the predictors for the k-NN method was also done on a hierarchical basis. To do so, 20 input combinations were considered in order to better investigate the contribution of various predictors and particularly DCB-Fe and DCB-Al in the predictive ability of the pattern-recognition method (Table 8–2).

Table 8-2. Different levels of input variables considered in the k-nearest neighbour algorithm to predict water retention of soils in Lower Congo.†

Level	Input attributes
1	Sand + Silt + Clay (SSC)
2	SSC + OC
3	SSC + pH
4	SSC + CEC
5	SSC + OC + pH
6	SSC + OC + CEC
7	SSC + pH + CEC
8	SSC + OC + pH + CEC
9	SSC + OC + pH + CEC + DCB-Fe
10	SSC + OC + pH + CEC + DCB-Al
11	SSC + OC + pH + CEC + DCB-Fe + DCB-Al
12	SSC + BD
13	SSC + BD + OC
14	SSC + BD + pH
15	SSC + BD + CEC
16	SSC + BD + pH + CEC
17	SSC + BD + OC + pH + CEC
18	SSC + BD + OC + pH + CEC + DCB-Fe
19	SSC + BD + OC + pH + CEC + DCB-Al
20	SSC + BD + OC + pH + CEC + DCB-Fe + DCB-Al

† SSC is sand, silt and clay, BD is bulk density, OC is organic carbon, pH is acidity, CEC is cation exchange capacity, DCB-Fe is dithionite-citrate-bicarbonate extractable iron, DCB-Al is dithionite-citrate-bicarbonate extractable aluminium.

8.2.3.3 Statistical criteria to evaluate the predictive ability of the MLR PTFs and the k-NN approach

Four statistical measures were selected to assess the predictive ability of the MLR PTFs and the k-NN algorithm at a given matric potential: the mean difference (MD), the mean absolute difference (MAD), the root mean square difference (RMSD) and the coefficient of determination (R^2).

$$MD = \frac{1}{N} \sum_{i=1}^N (\theta_{p_i} - \theta_{m_i}) \quad [8-1]$$

$$MAD = \frac{1}{N} \sum_{i=1}^N |\theta_{p_i} - \theta_{m_i}| \quad [8-2]$$

$$RMSD = \sqrt{\frac{1}{N} \sum_{i=1}^N (\theta_{p_i} - \theta_{m_i})^2} \quad [8-3]$$

$$R^2 = \frac{\left(\sum_{i=1}^N (\theta_{p_i} - \bar{\theta}_{p_i})(\theta_{m_i} - \bar{\theta}_{m_i}) \right)^2}{\sum_{i=1}^N (\theta_{p_i} - \bar{\theta}_{p_i})^2 \sum_{i=1}^N (\theta_{m_i} - \bar{\theta}_{m_i})^2} \quad [8-4]$$

where θ_{p_i} and θ_{m_i} are predicted and measured volumetric water contents for soil sample i ($\text{m}^3 \text{m}^{-3}$), respectively; $\bar{\theta}_{p_i}$ and $\bar{\theta}_{m_i}$ are the means of the predicted and measured volumetric water content, respectively; and N is the number of samples in the development or training soil dataset.

The overall prediction error in MLR PTFs is generally calculated using the RMSD index (Equation 8–3). In the calculation of RMSD, all members of the development dataset used to derive the equation are considered and therefore $N = 196$. In the k-NN approach, the Lower Congo dataset is used both as the training and the test dataset. Therefore, the RMSD is calculated based on a leave-one-out (LOO) procedure with $N = 196 - 1 = 195$. As the name suggests, the LOO method is performed by leaving only one sample out of the training dataset: all the samples except the one left out are used as a training dataset, and the model is validated on the sample left out (Mucherino et al., 2009).

8.3 Results and discussion

8.3.1 Hierarchical MLR PTFs to predict water retention

The correlation matrix between the physico-chemical properties of the Lower Congo soils and water content at different matric potentials is shown in Table 8–3.

Table 8-3. Pearson’s correlation matrix between the various potential predictors and water content at eight different matric potentials (N=196).

	Clay	Silt	Sand	BD	pH	OC	CEC	DCB_Fe	DCB_Al
θ_{-1} kPa	0.646**	0.098	-0.565**	-0.729**	-0.028	0.473**	0.513**	0.656**	0.682**
θ_{-3} kPa	0.731**	0.314**	-0.730**	-0.416**	-0.151*	0.462**	0.519**	0.713**	0.674**
θ_{-6} kPa	0.788**	0.494**	-0.856**	-0.323**	-0.239**	0.428**	0.621**	0.720**	0.674**
θ_{-10} kPa	0.798**	0.526**	-0.878**	-0.295**	-0.255**	0.421**	0.640**	0.721**	0.675**
θ_{-20} kPa	0.841**	0.558**	-0.927**	-0.267**	-0.255**	0.427**	0.658**	0.759**	0.701**
θ_{-33} kPa	0.833**	0.564**	-0.923**	-0.250**	-0.253**	0.411**	0.663**	0.752**	0.696**
θ_{-100} kPa	0.866**	0.516**	-0.928**	-0.270**	-0.235**	0.429**	0.646**	0.797**	0.729**
θ_{-1500} kPa	0.895**	0.366**	-0.885**	-0.268**	-0.187**	0.438**	0.585**	0.851**	0.766**

* shows significant correlation with volumetric water content at 0.05 significance level

** shows significant correlation with volumetric water content at 0.01 significance level

Based on different scenarios of availability of soil information, MLR PTFs were developed to predict water retention of soils at eight different matric potentials. Seven levels of availability were considered as depicted in Table 8-1.

The hierarchical PTFs that were derived by the stepwise procedure are of the following form:

$$\theta_{x \text{ kPa}} (\%) = a + b (\% \text{ sand}) + c (\% \text{ silt}) + d (\% \text{ clay}) + e (\text{BD, Mg m}^{-3}) + f (\text{pH, -}) + g (\% \text{ OC}) + h (\text{CEC, cmol}_c \text{ kg}^{-1} \text{ soil}) + i (\% \text{ DCB-Fe}) + j (\% \text{ DCB-Al}) \quad [8-5]$$

where $\theta_{x \text{ kPa}}$ is volumetric water content at a given matric potential, a is the intercept and $b, c, d, e, f, g, h, i,$ and j are the coefficients of the MLR PTF. The coefficients of Equation 8-5, obtained from the stepwise regression procedure are presented in Tables 8-4 to 8-10.

Table 8-4. Coefficients of pedotransfer functions to predict water retention at eight matric potentials developed based on information on sand, silt and clay content of 196 soils from Lower Congo.†

Predictors	Matric potential							
	-1 kPa	-3 kPa	-6 kPa	-10 kPa	-20 kPa	-33 kPa	-100 kPa	-1500 kPa
Intercept	37.234	37.785	44.196	43.520	42.302	41.929	34.137	21.046
Sand		-0.072	-0.252	-0.296	-0.344	-0.349	-0.283	-0.170
Silt								
Clay	0.139	0.093					0.071	0.175
BD								
pH								
OC								
CEC								
DCB-Fe								
DCB-Al								

† Soil properties that were retained as predictors in the stepwise regression procedure are in bold whereas soil properties that were assumed to be unknown are highlighted in grey.

Table 8-5. Coefficients of pedotransfer functions to predict water retention at eight matric potentials developed based on information on sand, silt, clay, pH and OC content of 196 soils from Lower Congo.†

Predictors	Matric potential							
	-1 kPa	-3 kPa	-6 kPa	-10 kPa	-20 kPa	-33 kPa	-100 kPa	-1500 kPa
Intercept	36.988	37.253	53.135	55.016	53.98	53.776	43.657	28.004
Sand		-0.068	-0.247	-0.289	-0.337	-0.343	-0.264	-0.156
Silt								
Clay	0.117	0.081					0.089	0.188
BD								
pH			-1.769	-2.275	-2.311	-2.345	-2.141	-1.564
OC	0.507	0.366						
CEC								
DCB-Fe								
DCB-Al								

† Soil properties that were retained as predictors in the stepwise regression procedure are in bold whereas soil properties that were assumed to be unknown are highlighted in grey.

Table 8-6. Coefficients of pedotransfer functions to predict water retention at eight matric potentials developed based on information on sand, silt, clay, pH, OC and CEC of 196 soils from Lower Congo.†

Predictors	Matric potential							
	–1 kPa	–3 kPa	–6 kPa	–10 kPa	–20 kPa	–33 kPa	–100 kPa	–1500 kPa
Intercept	31.168	37.253	50.374	51.633	50.995	50.293	40.565	28.004
Sand	0.053	-0.068	-0.225	-0.262	-0.313	-0.315	-0.236	-0.156
Silt								
Clay	0.156	0.081					0.097	0.188
BD								
pH			-1.801	-2.314	-2.346	-2.385	-2.201	-1.564
OC	0.339	0.366						
CEC	0.276		0.230	0.282	0.249	0.290	0.220	
DCB-Fe								
DCB-Al								

† Soil properties that were retained as predictors in the stepwise regression procedure are in bold whereas soil properties that were assumed to be unknown are highlighted in grey.

Table 8-7. Coefficients of pedotransfer functions to predict water retention at eight matric potentials developed based on information on sand, silt, clay, pH, OC, CEC, DCB-Fe and DCB-Al content of 196 soils from Lower Congo.†

Predictors	Matric potential							
	–1 kPa	–3 kPa	–6 kPa	–10 kPa	–20 kPa	–33 kPa	–100 kPa	–1500 kPa
Intercept	36.614	39.308	48.881	50.224	49.602	48.941	42.574	32.757
Sand		-0.085	-0.194	-0.234	-0.285	-0.288	-0.246	-0.196
Silt								
Clay								
BD								
pH			-2.141	-2.635	-2.663	-2.693	-2.312	-1.599
OC	0.610						-0.316	-0.302
CEC			0.185	0.239	0.207	0.249	0.218	
DCB-Fe		0.449					0.589	0.881
DCB-Al	3.849		1.813	1.711	1.691	1.642		

† Soil properties that were retained as predictors in the stepwise regression procedure are in bold whereas soil properties that were assumed to be unknown are highlighted in grey.

Table 8-8. Coefficients of pedotransfer functions to predict water retention at eight matric potentials developed based on information on sand, silt, clay and BD of 196 soils from Lower Congo.†

Predictors	Matric potential							
	-1 kPa	-3 kPa	-6 kPa	-10 kPa	-20 kPa	-33 kPa	-100 kPa	-1500 kPa
Intercept	67.228	48.080	44.196	43.520	42.302	41.929	26.478	8.405
Sand		-0.081	-0.252	-0.296	-0.344	-0.349	-0.276	-0.159
Silt								
Clay	0.089	0.067					0.091	0.207
BD	-20.057	-6.344					4.720	7.789
pH								
OC								
CEC								
DCB-Fe								
DCB-Al								

† Soil properties that were retained as predictors in the stepwise regression procedure are in bold whereas soil properties that were assumed to be unknown are highlighted in grey.

Table 8-9. Coefficients of pedotransfer functions to predict water retention at eight matric potentials developed based on information on sand, silt, clay, BD, pH, OC and CEC of 196 soils from Lower Congo.†

Predictors	Matric potential							
	-1 kPa	-3 kPa	-6 kPa	-10 kPa	-20 kPa	-33 kPa	-100 kPa	-1500 kPa
Intercept	65.958	45.902	50.374	51.633	50.995	40.877	27.944	9.775
Sand		-0.076	-0.225	-0.262	-0.313	-0.334	-0.218	-0.123
Silt						-0.067		
Clay	0.081	0.062					0.125	0.229
BD	-19.277	-5.256				7.196	6.467	8.821
pH			-1.801	-2.314	-2.346	-2.310	-2.014	-1.328
OC	0.216	0.283						
CEC			0.230	0.282	0.249	0.397	0.305	0.188
DCB-Fe								
DCB-Al								

† Soil properties that were retained as predictors in the stepwise regression procedure are in bold whereas soil properties that were assumed to be unknown are highlighted in grey.

Table 8-10. Coefficients of pedotransfer functions to predict water retention at eight matric potentials developed based on information on sand, silt, clay, BD, pH, OC, CEC, DCB-Fe and DCB-Al content of 196 soils from Lower Congo.†

Predictors	Matric potential							
	-1 kPa	-3 kPa	-6 kPa	-10 kPa	-20 kPa	-33 kPa	-100 kPa	-1500 kPa
Intercept	69.627	47.389	48.881	50.224	35.653	31.258	28.139	16.088
Sand	-0.027	-0.085	-0.194	-0.234	-0.277	-0.277	-0.242	-0.192
Silt								
Clay								
BD	-19.834	-5.477			7.847	9.947	8.361	9.711
pH			-2.141	-2.635	-2.448	-2.421	-2.108	-1.357
OC								
CEC			0.185	0.239	0.279	0.341	0.217	
DCB-Fe	0.333	0.378					0.404	0.698
DCB-Al			1.813	1.711	2.436	2.587	1.580	1.714

† Soil properties that were retained as predictors in the stepwise regression procedure are in bold.

All the hierarchical PTFs were compared based on the aforementioned statistical indices in order to identify the best equations as well as the best predictors of water retention for soils in Lower Congo. To simplify the comparison between each input level, the average RMSD was calculated across all the eight matric potentials (Table 8–11). Table 8–11 indicates that all the PTFs presented average RMSD values less than $0.05 \text{ m}^3 \text{ m}^{-3}$ with the highest average RMSD ($0.045 \text{ m}^3 \text{ m}^{-3}$) for Level 1 with only texture fractions as available input attributes and the lowest average RMSD ($0.040 \text{ m}^3 \text{ m}^{-3}$) for Level 7 where all the input soil properties were available for the stepwise regression process. Figure 8–1 illustrates a comparison between the measured and the predicted values of water retention by the MLR approach for all the eight matric potentials when level 7 of availability of input variables is considered. With the proposed point-based PTFs, rather good predictions can be obtained when only limited soil information is available. This means that these different PTFs can be applied by potential users to predict water retention of Lower Congo soils from the dry to the wet range with acceptable accuracy. Moreover, the potential user can select any SWRC model to estimate their parameters through curve-fitting (Šimůnek and Hopmans, 2002) and to subsequently predict water content at other potentials from the SWRC model.

It was also found that the predictive ability of the developed PTFs is much lower at the intermediate range of SWRC compared to the wet and the dry ranges. This observation is in agreement with the results found by Vereecken et al. (2010) and in Chapter 6.

Table 8-11. MAD, RMSD and R^2 values for hierarchical pedotransfer functions to predict water retention of soils in Lower Congo.

Level†	Input attributes	Avg_MAD	Avg_RMSD	Avg_ R^2
1	Sand + Silt + Clay (SSC)	0.034	0.045	0.742
2	SSC + pH + OC	0.033	0.044	0.756
3	SSC + pH + OC + CEC	0.032	0.043	0.765
4	SSC + pH + OC + CEC + DCB-Fe + DCB-Al	0.031	0.041	0.781
5	SSC + BD	0.032	0.044	0.778
6	SSC + BD + pH + OC + CEC	0.031	0.041	0.793
7	SSC + BD + pH + OC + CEC + DCB-Fe + DCB-Al	0.029	0.040	0.807

† SSC is sand (%), silt (%) and clay (%), BD is bulk density (Mg m^{-3}), OC is organic carbon (%), pH is acidity (-), CEC is cation exchange capacity ($\text{cmol}_c \text{ kg}^{-1}$ soil), DCB-Fe is dithionite-citrate-bicarbonate extractable iron (%), DCB-Al is dithionite-citrate-bicarbonate extractable aluminium (%), Avg_MAD is average mean absolute difference ($\text{m}^3 \text{ m}^{-3}$), Avg_RMSD is average root mean square difference ($\text{m}^3 \text{ m}^{-3}$), Avg_ R^2 is average coefficient of determination (-).

†† Levels 1, 2, 3, 4, 5, 6 and 7 represent selected levels of availability of soil information in the development (training) dataset.

When looking at all the MLR PTFs developed for Lower Congo soils, the relative importance of soil properties as (potential) predictors of water retention were found to differ in the wet, intermediate and dry ranges of the SWRC. Very close to saturation i.e. at -1 kPa, clay played an important role at all levels of predictor availability except when DCB-Fe and DCB-Al were present as potential predictors. In that case, they were selected in the stepwise process instead of clay content. At the wet range, BD was also an important predictor and seems to be replaced by OC when it is not available. Borgesen and Schaap (2005) found that adding BD and OM as predictors in their point-based PTFs for Danish soils increased their prediction capability. Sharma et al. (2006) found that using BD improved prediction of water retention at high matric potentials. In their study, Ghanbarian-Alavijeh and Millán (2010) observed that saturated water content was mainly affected by clay content and BD. At -3 kPa, sand becomes an important predictor along with clay and BD. On the other hand, OC still plays the substitute for BD, whereas DCB-Fe plays the substitute for clay content.

At the intermediate range of SWRC, i.e., from -6 to -33 kPa, the relative importance of (potential) predictors changes significantly. Clay content is not selected as potential predictor of water retention and is replaced by sand content. The same observation is made for BD, OC and DCB-Fe. Other soil properties, such as pH, CEC and DCB-Al, become the dominant predictors. It seems from these results that soil properties which are more related to structure, such as BD and DCB-Fe, are less important in the prediction of water retention at the intermediate range of the SWRC. As BD becomes less important, OC is also put aside in the selection process. This demonstrates the substitution role that OC may play when information on BD is not available. On the other hand, it also demonstrates that the role of OM in retaining water by adsorptive forces is very limited given the relatively low OM content in Lower Congo soils.

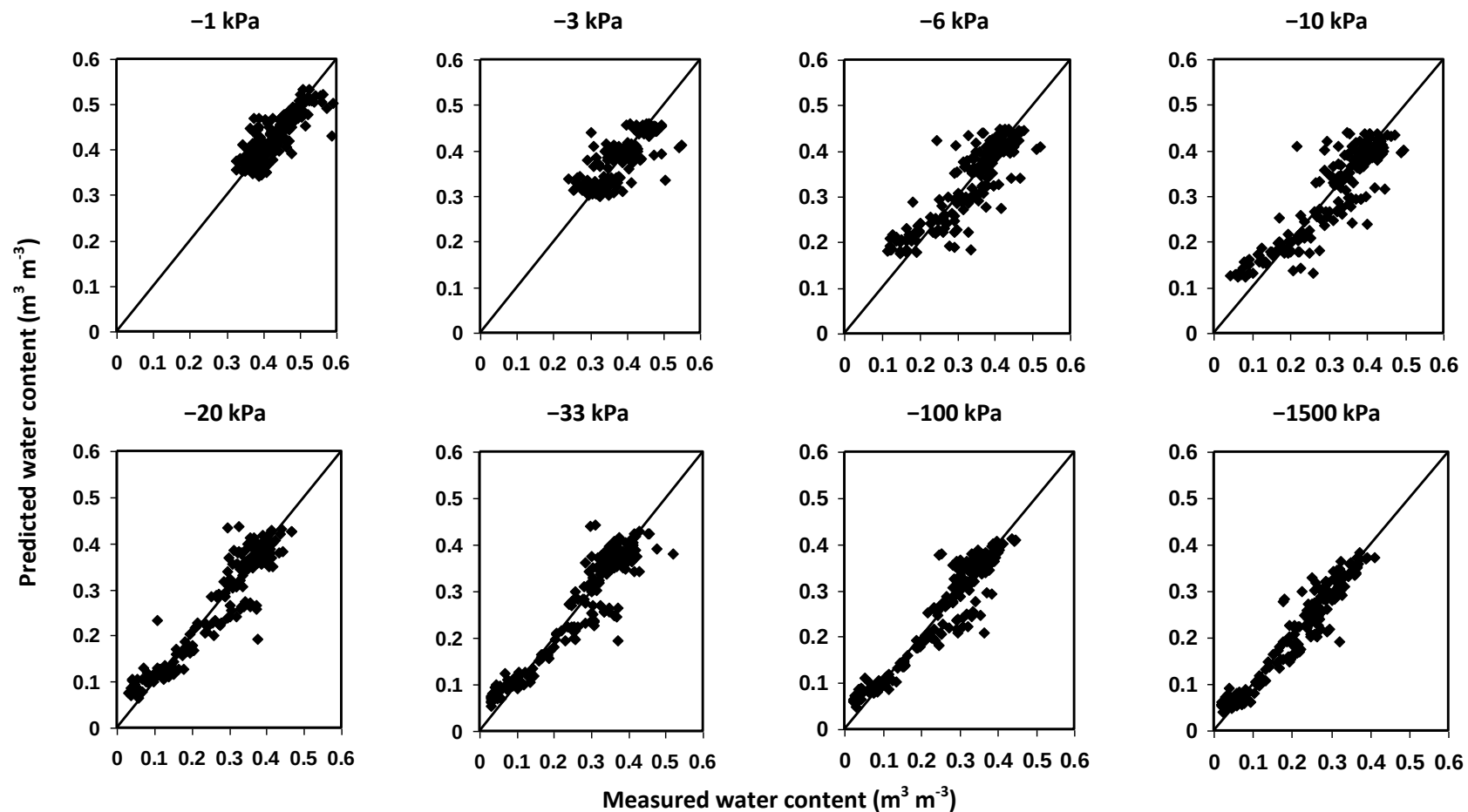


Figure 8–1. Measured versus predicted soil volumetric water contents at eight matric potentials of 196 soil samples from Lower Congo using the full set of available predictors in multiple linear regression PTFs.

In highly weathered soils like Lower Congo soils, properties like pH, CEC and DCB-Al are somewhat correlated. Cation exchange capacity is associated with electronegative colloidal substances such as clay minerals, colloidal silica and OM to a limited extent. Determination of CEC gives a good indication of the total amount of negative charges of the soils that are available for retention of cations or H^+ and Al^{+3} ions. The negative charges are due to the isomorphic substitution in layer silicate clays and are pH dependent on Fe and Al oxides and OM (Van Ranst et al., 1998, Baert et al., 1999, Yerima and Van Ranst, 2005). The results obtained in this study show that in the intermediate range, chemical soil properties play a more important role for water retention prediction than physical soil properties. In the dry range, most soil properties seem to play a role in the prediction of water retention particularly at -100 kPa, except silt content and OC. It seems that pH is preferred as predictor of water retention over OC in this case. Silt content should not be considered as a potential predictor of water retention of soils in Lower Congo. This can be explained notably by the low silt content of these soils as depicted in Figure 7–1.

On the other hand, DCB-Fe, DCB-Al and BD were found to improve the predictive ability of the MLR PTFs when included in the stepwise process. As previously mentioned in Chapter 5, these predictors, and in particular DCB-Fe and DCB-Al in case of many soils of the humid tropics, are closely associated to the structure of the soil and can therefore serve as proxies for soil structure. When BD is not available as a potential predictor, DCB-Fe and DCB-Al seem to provide a better substitute for BD as proxy to represent structure compared to OC. In our preliminary study reported in Chapter 5, it was observed that DCB-Fe appeared to yield a relatively better correlation with volumetric water content than BD for the examined soil samples. Schwertmann and Taylor (1989) wrote that pedochemically, Fe oxide is perhaps one of the most important components of ‘tropical’ soils, affecting many soil properties including CEC, aggregate stability and water retention among others. Moreover, in highly weathered soils such as Ferralsols and Nitisols with low OM/OC content, Fe and Al oxides play a prominent role as binding agents of negatively charged clay minerals, creating stable micro-aggregates within the size range of silt to fine sand (Schwertmann, 1988a,b; Arduino et al., 1989). This process leading to the formation of pseudosands, explains why the field texture of highly weathered soils is loamy rather than clayey as determined by texture analysis without pre-treatment in the lab (Cassel and Lal, 1992; Tomasella and Hodnett, 2004). In Chapter 7, it was shown that DCB-Fe and DCB-Al are among the best predictors of BD for soils in Lower Congo. These results are particularly encouraging for the prediction of water retention of soils in this region since BD is often lacking in various soil datasets in D.R. Congo. Bulk density was found to be an important predictor in the very wet range of the SWRC. In the very dry range, it was always selected as predictor in the stepwise process when available. On the other hand, BD seems to have less influence in the intermediate range compared to other soil properties.

8.3.2 PTFs based on pattern-recognition to predict water retention

A second approach, based on pattern-recognition, was tested to predict soil water retention at eight different matric potentials. This approach uses the k-NN variant developed and tested in Chapter 6. The summary statistics calculated through the eight matric potentials based on twenty different combinations of inputs are given in Table 8–12. It can be seen that the prediction performance of the k-NN technique is very good with the highest average RMSD value of $0.035 \text{ m}^3 \text{ m}^{-3}$ for SSC+pH and the lowest RMSD value of $0.031 \text{ m}^3 \text{ m}^{-3}$ for SSC + BD + OC + pH + CEC + DCBFe (+ DCB-Al). Though the difference in the average RMSD values may not always be statistically different between the various levels, we can observe that soil properties like pH seem to increase the overall prediction error, whereas BD, CEC, DCB-Fe and DCB-Al seem to improve the predictive ability of the k-NN

method. Using a pattern-recognition approach, the results show the relative importance of soil properties, such as DCB-Fe and DCB-Al, in the selection of the most similar soils to the target soil in relation with water retention. This shows once more the benefits of including DCB-Fe and DCB-Al as predictors when developing water retention PTFs for highly weathered soils in the humid tropics.

The results of this study show that the k-NN technique is a valuable tool for water retention prediction for soils of Lower Congo at all levels of availability of soil properties. The high predictive ability of this approach can be explained by the fact that soil instances in the Lower Congo dataset which are the closest to the target soil will get the highest weights in the calculation of the final estimate. Therefore, only the most relevant soils will be actually taken into consideration in the prediction of water retention at a given matric potential.

Table 8-12. MD, RMSD and R^2 values for the k-nearest neighbour algorithm with optimized settings and using 20 levels of input attributes.

Level [†]	Input attributes	Average MD	Average RMSD	Average R^2
		$m^3 m^{-3}$	$m^3 m^{-3}$	
1	Sand + Silt + Clay (SSC)	-0.001	0.034	0.715
2	SSC + OC	0.000	0.032	0.710
3	SSC + pH	-0.001	0.035	0.713
4	SSC + CEC	0.000	0.033	0.713
5	SSC + OC + pH	-0.002	0.034	0.714
6	SSC + OC + CEC	0.000	0.032	0.713
7	SSC + pH + CEC	-0.001	0.034	0.712
8	SSC + OC + pH + CEC	-0.001	0.033	0.715
9	SSC + OC + pH + CEC + DCBFe	-0.001	0.033	0.722
10	SSC + OC + pH + CEC + DCBAI	-0.002	0.033	0.718
11	SSC + OC + pH + CEC + DCB-Fe + DCB-Al	-0.002	0.033	0.720
12	SSC + BD	0.000	0.034	0.707
13	SSC + BD + OC	-0.001	0.033	0.711
14	SSC + BD + pH	0.000	0.034	0.708
15	SSC + BD + CEC	-0.001	0.032	0.708
16	SSC + BD + pH + CEC	-0.001	0.033	0.706
17	SSC + BD + OC + pH + CEC	0.000	0.032	0.712
18	SSC + BD + OC + pH + CEC + DCB-Fe	-0.001	0.031	0.721
19	SSC + BD + OC + pH + CEC + DCB-Al	-0.001	0.032	0.719
20	SSC + BD + OC + pH + CEC + DCB-Fe + DCB-Al	-0.001	0.031	0.722

[†] Levels 1 to 20 represent selected levels of availability of soil information in the training dataset.

8.3.3 Comparison between equation-based approach and pattern-recognition approach in prediction of water retention

From Table 8-13, it can be clearly seen that the k-NN approach outperforms the MLR approach in prediction of water retention at all the aforementioned seven input levels. As indicated previously, the k-NN or pattern-recognition approach presents the advantage of selecting the soils in the training dataset which are the closest to the target soil, whereas the MLR or equation-based approach derives a single equation out of the training (development) dataset. This means that if a larger dataset of soils from the humid tropics will be available, it will be more convenient to use a k-NN approach to predict water retention of the unknown soils than to redevelop each time new equation-based PTFs out of the enhanced dataset. As indicated in Chapter 6, the k-NN approach is much more flexible than the MLR approach in the sense that when more soil data are available, they can be included in the dataset

without the need of redeveloping new equations. In this study, applying the k-NN approach only necessitated to change the ‘training’ dataset (from the IGBP-Trop dataset to our own local Lower Congo dataset) when applying the algorithm presented in Chapter 6, without the need of reanalysing the data and developing new functions as with MLR. In a similar fashion, data from other regions could easily be added to, or replace those used in this study or in Chapter 6. However, a potential user which is not familiar with the pattern-recognition approach can still use the proposed MLR equations developed in this study as they yielded acceptable prediction performance for highly weathered soils of Lower Congo (average RMSD < 0.046 m³ m⁻³).

We also compared the RMSD values yielded by Lower Congo point-PTFs which use only texture fractions as potential predictors with the point-based PTFs recently developed by Minasny and Hartemink (2011) to predict water content at -10, -33 and -1500 kPa. As presented in Table 8-13, the highest RMSD values yielded by the Lower Congo PTFs using only texture in the prediction of water content at -10 kPa (0.053 m³ m⁻³), -33 kPa (0.048 m³ m⁻³) and -1500 kPa (0.037 m³ m⁻³) were sensitively lower than the RMSD values yielded by Minasny and Hartemink (2011) PTFs for predicting water content of Lower Congo soils at -10 kPa (0.065 m³ m⁻³), -33 kPa (0.068 m³ m⁻³) and -1500 kPa (0.048 m³ m⁻³).

Table 8-13. Prediction performance at various matric potentials in terms of accuracy of the MLR and the k-NN approaches developed for Lower Congo soils and in terms of reliability of the published PTFs of Hodnett and Tomasella (2002) and Minasny and Hartemink (2011) when applied on Lower Congo soils.

Pedotransfer functions (PTFs)		Matric potential							
		–1 kPa	–3 kPa	–6 kPa	–10 kPa	–20 kPa	–33 kPa	–100 kPa	–1500 kPa
Level†		RMSD (m ³ m ^{–3})							
1	MLR (SSC)	0.044	0.041	0.050	0.053	0.046	0.048	0.043	0.037
	k-NN (SSC)	0.033	0.032	0.036	0.037	0.033	0.035	0.032	0.030
2	MLR (SSC + OC + pH)	0.042	0.041	0.049	0.051	0.044	0.046	0.041	0.036
	k-NN (SSC + OC + pH)	0.031	0.033	0.041	0.040	0.034	0.032	0.030	0.028
3	MLR (SSC + OC + pH + CEC)	0.041	0.041	0.048	0.050	0.043	0.045	0.041	0.036
	k-NN (SSC + OC + pH + CEC)	0.031	0.034	0.040	0.039	0.034	0.032	0.030	0.027
4	MLR (SSC + OC + pH + CEC + DCB-Fe + DCB-Al)	0.039	0.040	0.047	0.049	0.042	0.044	0.038	0.032
	k-NN (SSC + OC + pH + CEC + DCB-Fe + DCB-Al)	0.031	0.033	0.040	0.039	0.033	0.030	0.028	0.026
5	MLR (SSC + BD)	0.033	0.040	0.050	0.053	0.046	0.048	0.043	0.036
	k-NN (SSC + BD)	0.031	0.035	0.041	0.040	0.032	0.033	0.030	0.027
6	MLR (SSC + BD + OC + pH + CEC)	0.032	0.040	0.048	0.050	0.043	0.044	0.040	0.034
	k-NN (SSC + BD + OC + pH + CEC)	0.030	0.034	0.040	0.040	0.032	0.030	0.027	0.025
7	MLR (SSC + BD + OC + pH + CEC + DCB-Fe + DCB-Al)	0.032	0.039	0.047	0.049	0.040	0.042	0.037	0.030
	k-NN (SSC + BD + OC + pH + CEC + DCB-Fe + DCB-Al)	0.030	0.034	0.039	0.039	0.030	0.028	0.025	0.023
		RMSD (m ³ m ^{–3})							
MLR-HT (SSC + BD + OC + pH + CEC)		0.040	0.059	0.058	0.057	0.052	0.050	0.043	0.039
MLR-MH (SSC + BD + OC)					0.065		0.068		0.048

† Levels 1, 2, 3, 4, 5, 6 and 7 represent selected levels of availability of soil information in the (development) training dataset. MLR is multiple linear regression, k-NN is k-nearest neighbour, MLR-HT is MLR PTFs developed by Hodnett and Tomasella (2002), MLR-MH is MLR PTFs developed by Minasny and Hartemink (2011), RMSD is root mean square difference (m³ m^{–3}), SSC is sand (%), silt (%) and clay (%), BD is bulk density (Mg m^{–3}), OC is organic carbon (%), pH is acidity (-), CEC is cation exchange capacity (cmol_c kg^{–1} soil), DCB-Fe is dithionite-citrate-bicarbonate extractable iron (%), DCB-Al is dithionite-citrate-bicarbonate extractable aluminium (%).

8.4 Conclusion

Pedotransfer functions to predict water content at various matric potentials of strongly weathered soils of Central Africa are missing. In order to respond to this need, two main approaches were used in this study: the equation-based approach and the pattern-recognition approach. In the first approach, different PTFs in the form of MLR equations were formulated to predict water content at eight different matric potentials based on seven different levels of availability of soil information. In the second approach, a variant of the k-NN algorithm developed in Chapter 6 was used with 20 different combinations of input attributes. Besides, the relative importance of soil properties like DCB-Fe and DCB-Al in the improvement of prediction of water retention in both approaches was demonstrated along with other more commonly used soil characteristics such as BD and CEC. Additionally, the predictive performance of the MLR PTFs and the k-NN PTFs using the same set of input variables was compared. The k-NN method outperformed the MLR method with an average RMSD $< 0.035 \text{ m}^3 \text{ m}^{-3}$. As indicated by Nemes et al. (2006a,b) and in Chapter 6, the k-NN method can be considered as a competitive alternative to more classical MLR PTFs due to the accuracy of the water retention estimation and its flexibility to incorporate new data without the burden to redevelop new equations. This is particularly advantageous in developing countries where soil databases for agricultural planning are at present sparse, though slowly developing.

Based on their needs and their familiarity towards one of the two approaches, potential users interested in predicting water retention of highly weathered soils of Lower Congo in particular and of Central Africa in general have the choice to use the proposed equations or to apply the k-NN algorithm using the present Lower Congo dataset as a training dataset. The derived PTFs can be used to provide estimates of water retention at FC and the PWP, or a complete SWRC. These hydraulic parameters are used as inputs in various simulation models or, e.g., to derive soil hydraulic properties maps. In future studies related to PTFs, other soil information like categorical soil structural data based on the profile descriptions should be investigated as potential predictors of water retention of soils in the humid tropics.

9 Performance of three pore-solid fractal approaches for soil water retention of highly weathered soils of the humid tropics

9.1 Introduction

Soil can retain water by means of the interactive capillary and adsorptive forces. Both forces result in what is called matric potential or matric suction Ψ (Hillel, 1998). The soil water retention curve (SWRC) describes the functional relationship between the matric potential of the soil Ψ and its gravimetric or volumetric water content W (Hillel, 1998; Jury and Horton, 2004). As an index of pore size distribution (POD), the SWRC is an important soil hydraulic characteristic. However, its complete mathematical description is required.

In Chapter 3, it was shown that most of the recent papers on prediction of water retention are dealing with soils from temperate areas with few papers dedicated to pedotransfer functions (PTFs) for soils in the humid tropics.

McBratney et al. (2002) stated that there are various ways to derive PTFs. According to one of these perspectives, PTFs can be developed by two approaches:

- a semi-physical approach: this approach attempts to describe a physical or chemical model relating easy-to-measure soil properties (e.g. texture) to more difficult-to-measure properties; and
- an empirical approach: this approach links easy-to-measure soil properties such as texture, organic carbon (OC), pH to the more difficult-to-measure soil properties by means of different numerical fitting methods.

In the semi-physical approach, hydraulic properties are derived based on physical attributes. Many semi-physical models are based on a representation of pore space geometry as a ‘bundle of cylindrical capillaries’ (BCC) (Millington and Quirk, 1961; Mualem, 1976). In the BCC model, it is implicitly assumed that certain pore sizes are completely filled by liquid whereas larger pores are completely empty at a given saturation level.

Semi-physical models recognize the shape similarity between the cumulative particle size distribution (PSD) curve and the SWRC. They offer valuable conceptual insights into the physical relations between the texture distribution and the POD. In our review study, we found that 91% of the published PTFs are based on the empirical approach and only 9% are based on the semi-physical approach. Among 35 PTFs dedicated to ‘tropical’ soils, only three of them were based on the semi-physical approach.

Pioneering work by Arya and Paris (1981) and Haverkamp and Parlange (1986) translated PSD data into a SWRC by means of the capillary equation. They assumed that the network of pores in the soil is a BCC. Tyler and Wheatcraft (1990) used fractal mathematics and scaled similarities to show that the empirical constant in the Arya and Paris (1981) model is equivalent to the fractal dimension of the tortuous fractal pore. Rieu and Sposito (1991a,b) presented a fully self-consistent mass fractal model of the soil water retention characteristics (Millán et al., 2006). The word ‘fractal’ as such has been coined by the mathematician Benoît Mandelbrot. The fractal dimension he described (Mandelbrot, 1983) is a measure of the irregularity degree of a given object seen at all scales (or resolutions) of observation, where the fractal structure is the one in which parts of it are similar in all. In simple words, a small piece of the object looks rather like a larger piece or the object as a whole. Therefore, the key property of fractal geometry is a degree of self-similarity across a range of spatial scales (or resolutions) of observation (Feder, 1988). Hydraulic PTFs based on the semi-physical approach require a detailed PSD (more than only clay, silt and sand content) and bulk density (BD) data. This may be one of the reasons of its limited application in developing PTFs especially for soils of the humid tropics.

In recent attempts to characterize soil water retention, one semi-physical model called the pore-solid fractal (PSF) model has gained increasing attention in the modelling community due notably to the growing interest for fractal theory in soil science. The interested reader can refer to Pachepsky et al. (2000) for more insights in the application of fractal models in soil science. In short, the PSF model is based on the assumption that both the PSD and the POD follow a power-law form with identical fractal dimensions (Perrier et al., 1999; Bird et al., 2000). In addition, the PSF model assumes that soil pores and soil particles should be identical in size and shape (Neimark, 1989).

Perrier et al. (1999) proposed a symmetric PSF model. The latter is characterized by the same geometric shape in the distribution of soil pores and the distribution of soil particles; both are assumed to be power functions with the same scaling component. Based on the PSF model, Bird et al. (2000) developed a new SWRC fractal model, which includes the Tyler and Wheatcraft (1990) and Rieu and Sposito (1991a,b) models as special cases. The symmetric PSF model of Bird et al. (2000) assumes that the fractal dimension ($D_{s,B}$) of the cumulative PSD could be directly substituted for the fractal dimension (D_p) of the SWRC. During recent decades, the relationship between the PSD model and the POD in soils has been the object of various studies. Researchers like Arya and Paris (1981) and Haverkamp and Parlange (1986) assumed that the relationship between PSD and POD is linear whereas others (e.g., Rouault and Assouline, 1998; Hwang and Powers, 2003) suggested that this relationship is non-linear.

Based on a concept of microscopic arrangement of different-sized soil particles, Hwang et al. (2011) modified the original PSF model of Bird et al. (2000) to develop a symmetry- and asymmetry-based PSF model. One of the objectives in modifying the original model was to determine whether or not differences exist in fractal dimensions between PSD and POD. The symmetry PSF model was formulated based on the Bird et al. (2000) original PSF model with the assumption of a linear relationship between PSD and POD (Haverkamp and Parlange, 1986) whereas the asymmetry PSF model was derived with the assumption of a non-linear relationship between PSD and POD (Rouault and Assouline, 1998). More mathematical details related to the original PSF model on the one hand and the modified asymmetry and symmetry PSF models on the other hand can be found respectively in Bird et al. (2000) and Hwang et al. (2011).

Hwang et al. (2011) demonstrated that the asymmetry-based PSF model performed better than the symmetry-based PSF model when applied to soils from the UNSODA database. These soils are mainly from Europe and North America (Nemes et al., 2001) and were sampled under temperate climates. Hwang et al. (2011) observed that the symmetry-based PSF model had a tendency to overestimate soil water content near saturation and to underestimate it in the drier moisture regime far from saturation. On the other hand, the asymmetry-based PSF model agreed well with observed water content. Overall they concluded that, except for coarse textured soils, the asymmetry PSF model performed better than the symmetry PSF model, indicating that the performance of the symmetry-based model is not stable over a wide range of soil textures because it does not address the effect of the microscopic arrangement of soil particles.

However, it is known that there are marked differences between highly weathered soils of the humid tropics and young soils of temperate regions in the relationships between texture, structure and water content (van den Berg et al., 1997; Tomasella et al., 2000; Hodnett and Tomasella, 2002). While ‘temperate’ soils containing more than 60% clay are considered as heavy clay soils with low permeability and a moderate to high available water capacity (AWC), highly weathered ‘tropical’ soils with clay contents ranging from 60% to 90% show high permeability and low AWC. These kaolinite dominated ‘tropical’ soils thus behave similarly to sandy soils in terms of permeability and AWC due

to their micro-aggregated structure (Mpulila et al., 2013), though they show high porosities typical for clay soils. Their stable microstructure is caused by the electrostatic forces between positively charged Fe and Al oxides and negatively charged kaolinite and organic matter (OM), with stable micro-aggregates being formed in the size range of silt to fine sand. This explains why their field texture is loamy rather than clayey as determined by laboratory analysis (Cassel and Lal, 1992). The strongly weathered soils of the humid tropics are generally low in OM content and are dominated by low-activity clay minerals, meaning that the effect of adsorption will be relatively low in favour of capillarity. As explained earlier, semi-physical fractal models are based on the theory of BCC which relies solely on capillarity, i.e., no adsorptive forces are directly considered. With their sandy hydraulic behaviour and their relatively low OM content, the PSF model may be suitable to describe and predict water retention for kaolinitic soils. Furthermore, since detailed PSD data (i.e., seven to eight fractions) are available for ‘tropical’ soils in well-known international databases such as IGBP-DIS (Tempel et al., 1996) and for soils previously or recently collected in Lower Congo with a large textural coverage (see Chapters 2 and 4), the question rises whether this information can be valorised for improving prediction of the SWRC.

The availability of detailed information on PSD and the very particular hydraulic behaviour of soils of the humid tropics have conducted us to investigate the applicability of the PSF approach to describe and predict water retention of highly weathered soils. To do so, this study aims at (1) validating the assumption that the asymmetry-based PSF model performs better than the symmetry-based PSF model when applied to describe water retention of highly weathered soils of the humid tropics; (2) comparing the ability of the PSF models and that of a well-known SWRC model to describe water retention for soils of the humid tropics; (3) assessing the prediction performance of the aforementioned PSF models to predict water retention at different matric potentials for soils of the humid tropics.

9.2 Materials and Methods

9.2.1 Study area: Lower Congo

The study area has been described in Chapter 2.

9.2.2 Soil datasets

In this study, a collection of 209 highly weathered soils from various tropical regions was used as calibration dataset to test the predictive ability of the three aforementioned PSF models, at first. These 209 soil samples are part of the IGBP-DIS international database from ISRIC (Tempel et al., 1996). The tropics are defined as the regions located between 25° N and 25° S and mainly under the (sub)-humid climates. Soils within the tropics but in temperate climates due to altitude or in dry areas are not included in the selected dataset. This ‘tropical’ dataset of 209 soils is referred to here as the IGBP-Trop-2 dataset. It is a subset of the IGBP-Trop dataset of 534 soils from the humid tropics that was used in Chapter 6. This subset contains essentially strongly weathered soils: Ferralsols (52.1%), Acrisols (29.7%), Nitisols (12.0%) and some sandy soils: Arenosols (6.2%) (IUSS Working Group WRB, 2006). Undisturbed and disturbed soil samples were collected under different land uses and at various depths. Tempel et al. (1996) provided the necessary references concerning the different analytical methods used to derive the soil physical and chemical properties recorded in the international IGBP-DIS database.

Secondly, a dataset of 196 soils from Lower Congo was used as both test and training dataset. These soils are mainly highly weathered soils of the humid tropics classified as Ferralsols, Acrisols and Nitisols (IUSS Working Group WRB, 2006), but other soil groups such as coarse textured soils i.e.

Arenosols (IUSS Working Group WRB, 2006) were also well represented. The 196 selected undisturbed soil samples are not part of the IGBP-DIS database, but were collected in 100 cm³ Kopecky rings under different land uses (savannah, forest, agricultural fields and old quarries) and at various depths in the soil profile during a sampling campaign in autumn of 2009. For this study, soils were selected in order to have the same WRB Reference Soil Groups in both the test and the training datasets when assessing the ability of the PSF approach to provide good predictions of water content in ‘tropical’ soils.

Soil properties selected for use in this study were the following: clay ($< 2 \mu\text{m}$), fine silt (2–20 μm), coarse silt (20–50 μm), very fine sand (50–100 μm), fine sand (100–250 μm), medium sand (250–500 μm), coarse sand (500–1000 μm), very coarse sand (1000–2000 μm) according to the USDA classification system (Soil Survey Staff, 1951), BD and retained (volumetric) water content (θ) at matric potentials: 0, –1, –3, –10, –20, –50, –250, –1500 kPa for the IGBP-Trop-2 dataset and –1, –3, –6, –10, –20, –33, –100, –1500 kPa for the Lower Congo dataset.

The physico-chemical characteristics of all soil samples (fine earth) from Lower Congo were determined using standard methods as described in Van Ranst et al. (1999). During these analyses, clay and silt fractions were determined by the pipette method of Köhn (1929) followed by a dry sieving to obtain the different sand fractions according to Gee and Bauder (1986). For soils in the IGBP-Trop-2 dataset, BD refers to the mass of a sample of undisturbed soil dried in the oven at 105 °C per unit bulk volume (ISSS Working Group, 1998, p.153). For Lower Congo soils, BD was determined by the core method following the procedures of Grossman and Reinsch (2002). After saturating the samples, they were gradually subjected to a tension of –10 kPa and subsequently weighted. Water content was then determined on a representative subsample and used to determine the sample’s oven-dry (105 °C) weight as detailed in Cornelis et al. (2005).

For the undisturbed samples from Lower Congo, the SWRC data pairs were determined from the wet to the dry range at various matric potentials. The hanging water-column method was used for matric potentials between –1 and –10 kPa using the sand box apparatus (Eijkelpamp Agrisearch Equipment, Giesbeek, the Netherlands), whereas for matric potentials between –20 and –1500 kPa, pressure chambers (Soil Moisture Equipment, Santa Barbara, CA) were used, following the procedures described in Cornelis et al. (2005). The coupled matric potential–water content pairs represent single measurements on single samples.

9.2.3 Grouping of soils per textural classes

In the Lower Congo dataset, while the predominant USDA textural classes were clay (47.4%), sand (13.8%), loamy sand (9.2%) and sandy clay loam (8.2%), the least represented classes were silt clay (1.5%), loam (1.0%) and sandy loam (7.1%). In the IGBP-Trop-2 dataset, the predominant textural classes were clay (49.8%), sandy clay loam (18.7%), sandy loam (8.6%) and sand (4.3%).

Soil textural classes were further grouped in three main categories according to FAO-UNESCO (1974): among the 196 soil samples of Lower Congo, 55.6% were fine textured (109 soil samples), 16.3% were medium textured (32 soil samples) and 28.1% were coarse textured (55 soil samples). In the IGBP-Trop-2 dataset, of 209 soil samples, 59.8% were fine textured (125 soil samples), 26.8% were medium textured (56 soil samples) and 13.4% were coarse textured (28 soil samples) as presented in Table 9–1. Soils were also categorized in unimodal and bimodal soils according to de Condappa et al. (2008) and as done in Chapter 4: among the 196 Lower Congo soil samples, 36.7% were unimodal (72 soil samples) and 63.3% were bimodal (124 soil samples). Less than half (45%) of

the soil samples in the IGBP-Trop-2 dataset were unimodal (94 soil samples) and the rest (55%) was bimodal (115 soil samples).

Table 9-1. Representation of the three aggregated FAO soil textural classes in the Lower Congo and the IGBP-Trop-2 datasets.

Dataset		Coarse textured	Medium textured	Fine textured	Total
Lower Congo	Number of soil samples	55	32	109	196
	Percentage, %	28.1	16.3	55.6	100
IGBP-Trop-2	Number of soil samples	28	56	125	209
	Percentage, %	13.4	26.8	59.8	100

9.2.4 Pore-solid fractal models used for fitting and prediction of soil water retention function

Three different PSF models were considered in this study: (1) the PSF model of Bird et al. (2000), (2) the symmetric PSF model of Hwang et al. (2011) and (3) the asymmetric PSF model of Hwang et al. (2011).

9.2.4.1 The PSF model of Bird et al. (2000)

Bird et al. (2000) provided a theoretical framework for fractal prediction of the SWRC from measurements of PSD. If soil structure follows PSF distributions, the SWRC is estimated directly from a power functional PSD under the assumption that both distributions have an identical fractal dimension.

If PSD of a soil is power-functional, its cumulative mass distribution is described as:

$$M(d_i) = L^3 \xi \rho_s \frac{s}{p+s} \left(\frac{d_i}{d_1} \right)^{3-D_{s,B}} \quad [9-1]$$

where $M(d_i)$ is the total mass of particle elements of a size smaller than or equal to d_i (the size of the i^{th} particle element, mm), L is the linear size of the initiator, ξ is a constant, $L^3 \xi$ is the bulk volume of PSF, ρ_s is the particle density, s and p are particle and pore fractions of PSF, respectively, d_1 is the size of the largest particle element and $D_{s,B}$ is the fractal dimension of cumulative mass distribution with B denoting the Bird model.

Substitution of d_i with d_1 into Equation (9-1) results in:

$$M(d_1) = L^3 \xi \rho_s \frac{s}{p+s} \quad [9-2]$$

Dividing Equation (9-1) by Equation (9-2) gives rise to the cumulative PSD function, $F(d_i)$, as:

$$F(d_i) = \frac{M(d_i)}{M(d_1)} = \left(\frac{d_i}{d_1} \right)^{3-D_{s,B}} \quad [9-3]$$

Equation (9-3) does not require the assumption of any average criterion for defining d_i (Millàn et al., 2003).

If POD of a soil follows the PSF distribution, SWRC can be described as:

$$S(h_i) = \frac{\theta(h_i)}{\theta_s} = \left(\frac{h_i}{h_1} \right)^{D_p-3} \quad [9-4]$$

$$S(h_i) = 1 \quad \text{for } h > h_b \quad [9-4b]$$

where h_i is the i^{th} matric head (cm), h_1 is the maximum matric head, h_b is bubbling pressure or air entry value (cm), $S(h_i)$ is the relative saturation at h_i , $\theta(h_i)$ is soil volumetric water content ($\text{cm}^3 \text{ cm}^{-3}$) at h_i , θ_s is the saturated water content and D_p is the fractal dimension of SWRC. The original symmetry-based PSF model assumes that both PSD and POD have an identical exponent of power-law function (Bird et al., 2000), i.e. $D_{s,B} = D_p$. Therefore, only two parameters ($D_{s,B}$ and h_1) are required to determine SWRC. Note that Equations (9-4) and (9-4b) represent a Campbell type of SWRC model (Campbell, 1974), which is based on the Brooks and Corey (1964) model with the residual water content set to zero.

9.2.4.2 Symmetry between PSD and POD: the symmetric PSF model of Hwang et al. (2011)

Symmetry between PSD and POD for a given soil means that the shape of POD is symmetric with respect to PSD. Therefore, POD can be superimposed onto PSD. To derive the symmetric PSF model, Hwang et al. (2011) formulated the original PSF model (Equation 9-4) with the assumption of a linear relationship between PSD and POD.

Haverkamp and Parlange (1986) suggested the symmetry between PSD and POD such that the pore radius of the i^{th} fraction (cm), r_i , is expressed as:

$$r_i = \frac{d_i}{\gamma} \quad [9-5]$$

where γ is the packing-related parameter of a soil (mm cm^{-1}), ranging generally from 0.1 to 100 (Hwang and Choi, 2006). Assuming perfectly wettable surfaces ($\cos\theta = 1$), the matric head, h_i (cm), is associated with r_i (cm) using the simplified capillary equation (Brutsaert, 1966):

$$h_i = \frac{0.149}{r_i} \quad [9-6]$$

$S(h_i)$ can be expressed as a function of d_i , d_1 and D_p by substituting Equations (9-5) and (9-6) into Equation (9-4):

$$S(h_i(d_i)) = \left(\frac{h_i(d_i)}{h_1(d_1)} \right)^{D_p-3} = \left(\frac{d_i}{d_1} \right)^{3-D_p} \quad [9-7]$$

where

$$h_i(d_i) = \frac{0.149\gamma}{d_i} \quad [9-7b]$$

Because the PSF model assumes that both PSD and POD have an identical exponent of the power-law function, D_p on the right-hand side of Equation (9-7) can be replaced by the fractal dimension of the cumulative mass distribution $D_{s,H}$ (with H denoting the Hwang model):

$$S(h_i(d_i)) = \left(\frac{d_i}{d_1} \right)^{3-D_p} = \left(\frac{d_i}{d_1} \right)^{3-D_{s,H}} \quad [9-8]$$

Equation (9–8) shows that SWRC can be described explicitly by the cumulative PSD function (Equation 9–3), indicating that the original PSF model (Equation 9–4) is based on the symmetry between PSD and POD. Hwang et al. (2011) defined Equations (9–7) and (9–8) as the symmetry-based PSF model, with two parameters to be determined, $D_{s,H}$ and d_1 .

9.2.4.3 Asymmetry between PSD and POD: the asymmetric PSF model of Hwang et al. (2011)

If for a given soil the shape of POD is asymmetric to that of PSD, POD cannot be superimposed onto PSD. To derive the asymmetric PSF model, Hwang et al. (2011) formulated the original PSF model (Equation 9–4) with the assumption of a non-linear relationship between PSD and POD.

Rouault and Assouline (1998) suggested the asymmetry relationship between PSD and POD such that r_i is related to d_i by:

$$r_i = \left(\frac{d_i}{u} \right)^{\frac{1}{\nu}} \quad [9-9]$$

where u is the packing characteristic of a soil (mm cm^{-1}) and its physical meaning and range are similar to those of the parameter γ (Hwang and Choi, 2006). The parameter ν relates to respective shapes of the particles and corresponding pores, ranging generally from 0.2 to 2.0 (Hwang and Choi, 2006) and affecting the shape of POD.

$S(h_i)$ can be expressed as a function of d_i , d_1 , D_p and ν by substituting Equations (9–6) and (9–9) into Equation (9–4):

$$S(h_i(d_i)) = \left(\frac{h_i(d_i)}{h_1(d_1)} \right)^{D_p-3} = \left(\frac{d_i}{d_1} \right)^{\frac{3-D_p}{\nu}} \quad [9-10]$$

where

$$h_i(d_i) = \frac{0.149u^{\frac{1}{\nu}}}{d_i^{\frac{1}{\nu}}} \quad [9-10b]$$

Because the PSF model assumes that both PSD and POD have an identical power-law exponent, the power on the right-hand side of Equation (9–10) is equal to that of the cumulative PSD function (Equation 9–3).

Equating Equation (9–3) with Equation (9–10) results in:

$$S(h_i(d_i)) = \left(\frac{d_i}{d_1} \right)^{\frac{3-D_p}{\nu}} = \left(\frac{d_i}{d_1} \right)^{3-D_{s,H}} \quad [9-11]$$

with the relationship of:

$$D_p = \nu(D_{s,H} - 3) + 3 \quad [9-12]$$

Equation (9–12) is referred to as the general relationship between $D_{s,H}$ and D_p (when $\nu = 1$, then $D_p = D_{s,H}$). Herein, Hwang et al. (2011) define Equations (9–10) and (9–12) as the asymmetry PSF approach by which a SWRC is estimated from the cumulative PSD with defined ν , $D_{s,H}$ and h_1 .

Because d_1 is the size of the largest particle in the original PSF model (Bird et al., 2000), there is no need to define $d_1 = 2 \text{ mm}$ (the largest mineral particle by definition in case of conventional texture analysis). Therefore, d_1 was set by Hwang et al. (2011) as a fitting parameter for both the symmetry- and asymmetry-based PSF models. However, the maximum matric head (h_1) is required to estimate

the SWRC (Equation 9–4). While the symmetry-based PSF model (Equation 9–13) estimated h_l by fitting the SWRC data with two parameters (γ and d_l), the asymmetry PSF model (Equation 9–14) proposed herein estimated h_l by fitting the SWRC data with three parameters (u , d_l and v):

$$h_l(d_l) = \frac{0.149\gamma}{d_l} \quad [9-13]$$

and

$$h_l(d_l) = \frac{0.149u^{1/v}}{d_l^{1/v}} \quad [9-14]$$

Although both models (Equations 9–13 and 9–14) are formulated on the basis of the PSF distribution using the same PSD data, formulation of the SWRC is different depending on the assumption of symmetry or asymmetry between PSD and POD. This results in a different number of parameters required for the calculation of D_p and h_l . In other words, unlike the symmetry-based PSF model that requires h_l and $D_{s,H}$ to calculate D_p , the asymmetry-based PSF model requires h_l , $D_{s,H}$ and v .

9.2.4.4 Fitting procedure

Four parameters, namely D_s , d_l (for the PSD), D_p and h_l (for the SWRC), were estimated for each Lower Congo soil by fitting Equations (9–3) and (9–4) to the observed PSD and water retention data, respectively by an iterative non-linear optimization procedure using Mathematica 7.0 (Wolfram Research, Inc., Champaign, IL).

For the symmetry-based PSF model, the parameter γ in Equation (9–13) was calibrated for each FAO aggregated soil textural class, using minimization of the following objective function:

$$O(\gamma) = \sum_{i=1}^{N_r} \left\{ \frac{0.149\gamma}{d_{l,i}} - h_{l,i} \right\}^2 \quad [9-15]$$

where $h_{l,i}$ is the fitted maximum matric head, $d_{l,i}$ is the fitted diameter of particles, and N_r is the number of soils within each aggregated FAO textural class of the (calibration) dataset.

For the asymmetry PSF model, the parameter v in Equation (9–12) was optimized for each FAO textural class in the calibration dataset, using minimization of the following objective function:

$$O(v) = \sum_{i=1}^{N_r} \left(v(D_{sH,i} - 3) + 3 - D_{p,i} \right)^2 \quad [9-16]$$

where $D_{sH,i}$ and $D_{p,i}$ are the fitted fractal dimensions for the PSD and SWRC calibration dataset, respectively.

After optimization of the variable v , the other parameter u in Equation (9–14) was optimized for each FAO textural class in the calibration dataset, using minimization of the following objective function:

$$O(u) = \sum_{i=1}^{N_r} \left\{ \frac{0.149u^{1/v}}{d_{l,i}^{1/v}} - h_{l,i} \right\}^2 \quad [9-17]$$

9.2.4.5 Comparison of the fitting performance of the PSF models with the van Genuchten (1980) model

The ability of the PSF models to describe water retention function was compared with the one of the van Genuchten (1980) model. The latter has shown good results for a variety of soils and is considered

as a reference in its ability to characterize water retention condition in unsaturated soil. The van Genuchten (1980) model was fitted to the observed water retention data with the aforementioned non-linear optimization procedure using Mathematica 7.0. Eventually, water content values were read at particular matric potentials off the fitted curve and compared to the measured values.

9.2.4.6 Prediction procedure

After D_s and d_l were estimated from the PSD data in the Lower Congo dataset using the non-linear regression procedure, the predictive ability of the PSF models can be tested with coupled matric potential–water content data from the IGBP-Trop-2 calibration dataset and using the calibrated parameters γ , u and v for each FAO-aggregated textural class. The IGBP-Trop-2 calibrated parameters γ , u and v were used to derive water content at the predefined matric potentials.

9.2.4.7 Statistical criteria to evaluate the predictive ability of the three PSF models

Three statistical measures were selected to compare the fitting ability of these semi-physical models with the van Genuchten (1980) model and to assess the predictive ability of the PSF models at a given matric potential: the mean difference (MD), the root mean square difference (RMSD) and the coefficient of determination (R^2).

$$MD = \frac{1}{N_t} \sum_{i=1}^{N_t} (\theta_{p_i} - \theta_{m_i}) \quad [9-18]$$

$$RMSD = \sqrt{\frac{1}{N_t} \sum_{i=1}^{N_t} (\theta_{p_i} - \theta_{m_i})^2} \quad [9-19]$$

$$R^2 = \frac{\left(\sum_{i=1}^{N_t} (\theta_{p_i} - \bar{\theta}_{p_i})(\theta_{m_i} - \bar{\theta}_{m_i}) \right)^2}{\sum_{i=1}^{N_t} (\theta_{p_i} - \bar{\theta}_{p_i})^2 \sum_{i=1}^{N_t} (\theta_{m_i} - \bar{\theta}_{m_i})^2} \quad [9-20]$$

where θ_{p_i} and θ_{m_i} are predicted and measured volumetric water contents for soil sample i ($\text{m}^3 \text{m}^{-3}$), respectively; $\bar{\theta}_{p_i}$ and $\bar{\theta}_{m_i}$ are the means of the predicted and measured volumetric water content, respectively; and N_t is the number of soil samples in the test dataset.

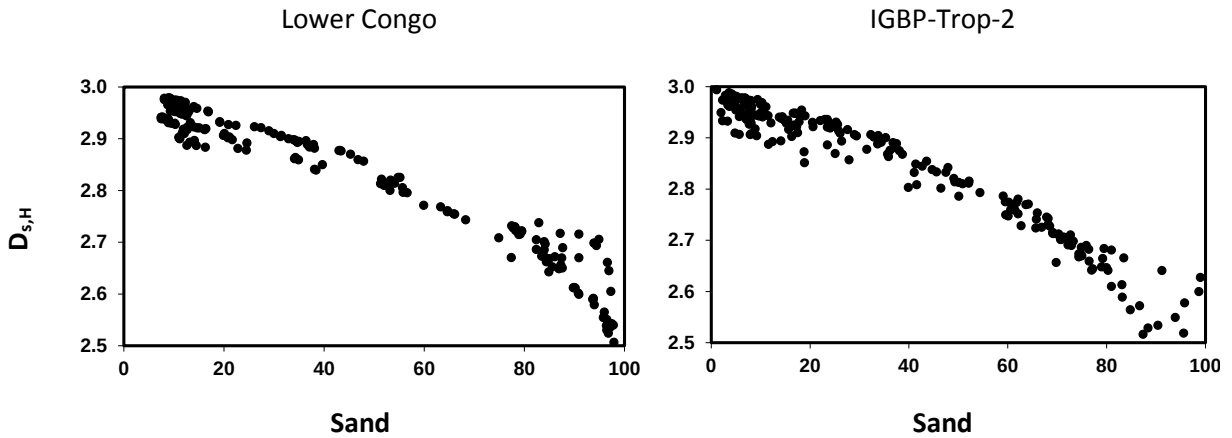
9.3 Results and discussion

9.3.1 Relationships between the fractal dimensions of PSD and SWRC

Tables 9–2 and 9–3 present the fitting parameter values for the cumulative PSD and the SWRC computed for three aggregated FAO textural classes for 196 soil samples of the Lower Congo dataset and for 209 soil samples of the IGBP-Trop-2 dataset. From Tables 9–2 and 9–3, one can observe that the value of the fractal dimension of the PSD curve derived from the original PSF model of Bird et al. (2000) (designated here as $D_{s,B}$) or the modified symmetric and asymmetric models of Hwang et al. (2011) (designated here as $D_{s,H}$) increased from coarse to fine textured soils. According to Perfect and Kay (1995), an increase of the fractal dimension of the PSD curve with clay content indicates that the pore space is filled more with finer particles in the PSF model. Various researchers (Tyler and Wheatcraft, 1992; Bittelli et al., 1999; Huang and Zhang, 2005) also reported a strong dependence of the fractal dimension of PSD on texture for temperate-climate soils. The standard deviations of the $D_{s,B}$, $D_{s,H}$ and D_p values were generally greater for the coarse textured soils compared to the fine textured soils probably due to a larger variation in PSD and SWRC of the coarse textured soils.

The highly weathered soils of Lower Congo and the IGBP-Trop-2 dataset show a similar curvilinear decrease of fractal dimension of the PSD curve with sand content in form of a power function as depicted in Figure 9–1(a). It was also observed that the value of the fractal dimension of SWRC (D_p) increased with clay content due to the formation of finer pore spaces as indicated by Perfect and Kay (1995). On the other hand, a strong power function ($R^2 = 0.854$) was found to better represent the relationship between clay and the fractal dimension of SWRC (D_p) for the Lower Congo dataset whereas this was not the case for the IGBP-Trop-2 dataset ($R^2 = 0.335$) as can be seen from Figure 9–1(b). This means that the nature of the relationship between the fractal dimension of the SWRC and clay content was different in both datasets. This can be an indication that the (fractal) geometry of POD of most soils is different in the two datasets.

(a)



(b)

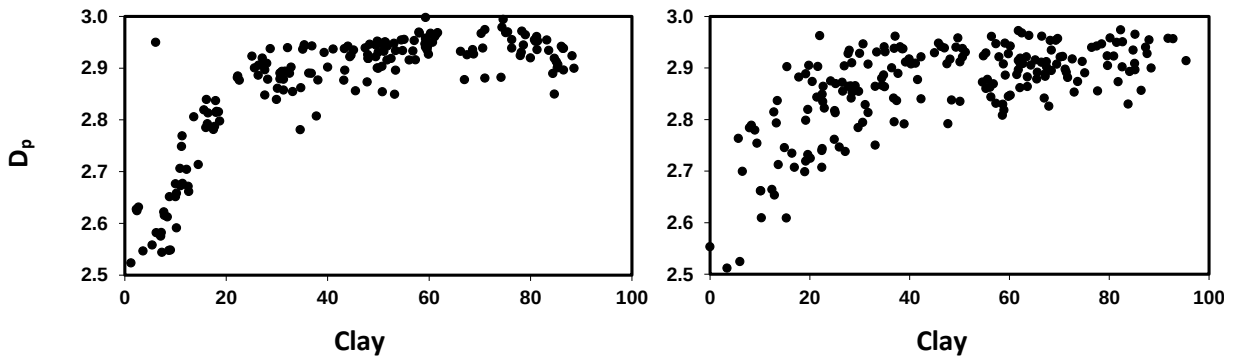


Figure 9–1. (a) Fractal dimension of particle size distribution curve versus sand content and (b) fractal dimension of soil water retention curve versus clay content for the pore-solid fractal model of Hwang et al. (2011).

The following power function was found between clay content and D_p for the Lower Congo dataset:

$$D_p = 2.396 \times \text{clay}^{0.05} \quad [9-21]$$

The following power function could be established for the IGBP-Trop-2 dataset:

$$D_p = 2.683 \times \text{clay}^{0.02} \quad [9-22]$$

Since clay content explains the variation in D_p with more than 85% with Equation (9–21), it can be used as the only predictor variable to derive the fractal dimension of SWRC of Lower Congo soils. However, more investigations need to be conducted to validate the observed strong power relation

between D_p and clay content. On the other hand, the poor power relationship found in Equation 9–22 may be partially due to the fact that PSD and SWRC data from the IGBP-Trop-2 international dataset originated from various sources.

Table 9-2. Fitting parameter values for the cumulative PSD and the SWRC per FAO textural class for 196 soil samples of the Lower Congo dataset.†

FAO textural class	Nbr of soils	$D_{s,B}$ (-)	h_{min} (kPa)	$D_{s,H}$ (-)	d_1 (mm)	D_p (-)	h_l (kPa)	γ (mm cm ⁻¹)	ν (-)	u (mm cm ⁻¹)
Coarse textured	55	2.525 (0.136)†	0.996 (0.376)	2.645 (0.065)	1.296 (0.096)	2.615 (0.137)	0.765 (0.310)	66.4	1.105	100.7
Medium textured	32	2.793 (0.042)	1.313 (1.174)	2.808 (0.055)	1.224 (0.297)	2.877 (0.039)	0.488 (0.497)	40.3	0.622	10.0
Fine textured	109	2.921 (0.040)	1.618 (2.172)	2.926 (0.039)	1.278 (0.484)	2.931 (0.032)	0.893 (1.454)	64.3	0.805	27.6

† Values between parentheses represent standard deviations

Table 9-3. Fitting parameters values for the cumulative PSD and the SWRC per FAO textural class for 209 soil samples of the IGBP-Trop-2 dataset.†

FAO textural class	Nbr of soils	$D_{s,B}$ (-)	h_{min} (kPa)	$D_{s,H}$ (-)	d_1 (mm)	D_p (-)	h_l (kPa)	γ (mm cm ⁻¹)	ν (-)	u (mm cm ⁻¹)
Coarse textured	28	2.523 (0.202)†	0.851 (0.339)	2.618 (0.066)	1.388 (0.150)	2.637 (0.158)	0.639 (0.277)	59.9	0.978	55.1
Medium textured	56	2.769 (0.043)	0.944 (0.944)	2.756 (0.066)	1.329 (0.277)	2.842 (0.067)	0.421 (0.513)	40.5	0.634	11.2
Fine textured	125	2.923 (0.042)	1.100 (3.281)	2.926 (0.045)	1.131 (0.428)	2.904 (0.043)	0.976 (3.115)	46.3	0.996	45.5

† Values between parentheses represent standard deviations

9.3.2 Fitting performance of three pore-solid fractal models tested on 196 soil samples of the Lower Congo dataset

The ability of the PSF model of Bird et al. (2000), and the symmetric and asymmetric PSF models of Hwang et al. (2011) to describe the soil water retention at eight different matric potentials for 196 soils of the Lower Congo dataset was tested and the results in terms of RMSD are given in Table 9–4. The fitting performances of the PSF models were better for fine and medium textured soils than for coarse textured soils. This may be due to the lack of stable structure in sandy soils which can impede a good adjustment of packing-related parameters during the non-linear optimization process. The SWRC model of van Genuchten (1980) performed better than the PSF models for coarse textured soils. The symmetric PSF model of Hwang et al. (2011) yielded comparable fitting performance to the van Genuchten (1980) model for fine and medium textured soils.

It can be seen that the symmetric PSF model of Hwang et al. (2011) fits the observed data better than the original PSF model of Bird et al. (2000). On the other hand, the asymmetric PSF model of Hwang et al. (2011) shows poor fitting performance for Lower Congo soils. Cihan et al. (2009) indicated that most fractal models do not include an explicit description of incomplete pore connectivity, which can result in incomplete drainage of pores as matric potential is decreased. This may partially explain why for coarse textured soils, estimates of physically-based parameters, such as the mass fractal dimension, obtained by fitting PSF models to experimental water retention data, may not be accurate.

Based on the above results, we can say that when detailed information of the PSD curve is available, the symmetric PSF model of Hwang et al. (2011) can be a competitive alternative to the van Genuchten (1980) equation to describe the SWRC of medium and fine textured soils. The symmetric PSF model may be useful when one wants to harmonize datasets from various origins (e.g., for developing PTFs) and provide water contents at the same matric potentials, whereas now, SWRC databases are based on different matric potentials. Nowadays, this is typically done by fitting a curve to the data and using a closed-form equation (e.g., Brooks and Corey, 1964; Campbell, 1974; van Genuchten, 1980) to derive the water content at common matric potentials off the fitted curve. With that approach, one would be able to estimate the water content at the target matric potential in a relatively direct fashion.

Furthermore, these results show that for Lower Congo soils, the assumption of symmetry between PSD and POD yields better results than the assumption of asymmetry between them. The same observations were also made for soils of the IGBP-Trop-2 dataset. Hwang et al. (2011) found somewhat opposite results with temperate-climate soils of the UNSODA database particularly for fine textured soils. Further studies are needed in order to establish whether these opposite results are related to marked differences in physico-chemical, structural and mineralogical properties between ‘temperate’ and ‘tropical’ soils.

Table 9-4. Performance in terms of RMSD values of the van Genuchten (1980) and three pore-solid fractal models to describe the soil water retention curve for 196 soils of the Lower Congo dataset.†

Model for SWRC	Matric potential							
	-1 kPa	-3 kPa	-6 kPa	-10 kPa	-20 kPa	-33 kPa	-100 kPa	-1500 kPa
	RMSD (m ³ m ⁻³)							
	Coarse textured soils							
Van Genuchten	0.005	0.009	0.010	0.011	0.011	0.007	0.008	0.013
PSF Bird	0.065	0.073	0.042	0.039	0.033	0.034	0.046	0.061
PSF symmetric	0.025	0.064	0.024	0.026	0.035	0.026	0.014	0.030
PSF asymmetric	0.056	0.085	0.053	0.057	0.056	0.053	0.051	0.059
Medium textured soils								
Van Genuchten	0.030	0.024	0.018	0.016	0.025	0.021	0.019	0.051
PSF Bird	0.081	0.021	0.024	0.031	0.031	0.048	0.056	0.078
PSF symmetric	0.020	0.010	0.022	0.021	0.022	0.015	0.010	0.025
PSF asymmetric	0.052	0.066	0.079	0.080	0.075	0.069	0.063	0.065
Fine textured soils								
Van Genuchten	0.014	0.022	0.020	0.019	0.019	0.019	0.018	0.043
PSF Bird	0.063	0.028	0.016	0.019	0.029	0.034	0.035	0.047
PSF symmetric	0.021	0.012	0.015	0.019	0.026	0.021	0.012	0.025
PSF asymmetric	0.039	0.059	0.075	0.083	0.091	0.092	0.089	0.088

† SWRC is soil water retention curve; van Genuchten is the SWRC model of van Genuchten (1980); PSF Bird is the original pore-solid fractal (PSF) model of Bird et al. (2000); PSF symmetric and PSF asymmetric are the modified PSF models of Hwang et al. (2011); RMSD is root mean square difference.

9.3.3 Prediction performance of the symmetric PSF model calibrated on the IGBP-Trop-2 dataset and tested on Lower Congo soils

Taking into account its very good fitting ability, the symmetric PSF model of Hwang et al. (2011) was selected and further tested as a potential predictive model of water retention at different matric potentials. Using this PSF model for prediction purposes requires the calibration of some of its parameters based on a training dataset and the application of the calibrated parameters values on a test dataset for validation. The parameter h_1 is related to the parameter γ and d_1 by Equation 9–7b. Therefore, the packing-related parameter γ is the only parameter that needs to be optimized (calibrated) based on SWRCs of the considered training dataset (i.e. IGBP-Trop-2).

Hwang et al. (2011) optimized the parameter γ for five aggregated USDA classes (Soil Survey Staff, 1993): fine, moderately fine, medium, moderately coarse, and coarse. In this study, due to a limited number of soils in some USDA aggregated textural classes for the Lower Congo dataset as well as for the IGBP-Trop-2 dataset (e.g., loam, sandy clay, silty clay and silty clay loam), a calibration by three aggregated FAO textural classes was found to be the best option. These aggregated FAO textural classes, i.e., fine, medium, and coarse (FAO-UNESCO, 1974) were previously used in Chapter 4 dealing with the fitting performance of various PSD models when applied to soils of the humid tropics. The parameter γ in Equation 9–7b was calibrated for each FAO soil textural class using minimization of an objective function (see Equation 9–15).

The prediction performance in terms of bias (MD), overall error (RMSD) and goodness-of-fit (R^2) of the symmetric PSF model calibrated on the IGBP-Trop-2 dataset when tested for Lower Congo soils is presented in Table 9–5. When compared to the hierarchical PTFs based on MLR and on the k-NN

approach on the same dataset (see Chapter 8), these results show somewhat limited prediction performance of the calibrated PSF model particularly for fine textured soils. This may suggest that the performance of the symmetry-based PSF model is not stable over a wide range of soil textures. However, this should be confirmed by further studies with e.g., more detailed textural classes.

The better prediction of water content for medium textured soils compared to fine textured soils based on the PSF model (see also Figure 9–2) may be partially due to stable micro-aggregates which are known to be formed in the size range of silt to fine sand of highly weathered tropical soils, i.e., hence being more dominant in medium textured soils. This might have contributed to a better calibration of the packing-related parameter γ . No particular differences were found between unimodal and bimodal soils with regard to the prediction performance of the symmetric PSF model. On the other hand, there is tendency of the PSF model to overestimate water content for fine textured soils with high clay content and to underestimate water content for coarse textured soils with high sand content. This may be attributed to differences in some soil properties between the calibration or training dataset (IGBP-Trop-2) and the validation dataset (Lower Congo) that may hinder a good calibration of the packing parameter γ . As indicated by Hwang et al. (2011), packing involves multi-sized particle arrangements and smaller particles can fit within the pore pockets formed by the packing of larger particles. On the other hand, Hwang et al. (2011) observed that their modified PSF model cannot interpret pore interactions and/or micro-aggregation.

Previously, Basile and D'Urso (1997) indicated that the complexity of the porous system may be of particular relevance for fine textured soils, where the geometry of pores is very irregular. This has been considered by Tuller et al. (1999) as one of the shortcomings of the BCC theory in which the cylindrical pore geometry is assumed. Tuller et al. (1999) wrote that because many types of porous media are formed by aggregation of primary particles and various (often flat) mineral surfaces, the resulting pore space is more realistically described by angular or slit-shaped pore cross sections rather than by cylindrical capillaries as also indicated by previous researchers (e.g., Li and Wardlaw, 1986; Mason and Morrow, 1991).

Some factors like content and/or nature of the OM present in a soil sample could also have an influence on the geometry of the pores and therefore on the shape of the SWRC. Therefore, large differences in content and/or nature of OM between soils in the training and the validation datasets may lead to very different packing-related parameter values which can also lead to poor prediction performance after calibration. In this study, average OC content of the test dataset was found to be twice as much as the calibration dataset as illustrated in Figure 9–3. Additionally, OM content can provide more adsorptive forces which can retain more water and therefore increase the water retention capacity especially in sandy soils (Hillel, 1998). Williams et al. (1983) observed that the presence of montmorillonite in the studied soil samples was strongly associated with their water retention characteristics. On the other hand, free Fe oxide, present in half of the soils examined by Williams et al. (1983), was also found to strongly influence soil water retention.

Though in a limited extent, low-activity clay minerals such as kaolinite can contribute to adsorptive forces in partially saturated soil. Comparing the hydrophilic/hydrophobic character of several samples of kaolinites (dominant in 'tropical' soils) and illites (dominant in 'temperate' soils), Saada et al. (1995) showed that the nature and the genesis condition of the clays affect the adsorption of water. Moreover, great differences were found between the specific hydrophilicity values of the two clay minerals. They observed that only 25% of the kaolinite surface was hydrophilic, whereas this value extends to 40% for illite. In fine and medium textured soils, the presence of relatively large exposed

surface areas and related adsorptive forces are not well taken into account by most physical conceptual models.

In proposing a new conceptual framework for liquid distribution in partially saturated porous media, Tuller et al. (1999) mentioned that in practice, the matric potential is attributed to capillary forces only as evidenced from the determination of media POD from soil water characteristics measurements which relies solely on cylindrical capillarity, ignoring the role of surface area and adsorbed liquid films. Hence, based on the BCC theory, soil water is assumed to be held by capillary forces only in ‘surfaceless’ cylindrical pores of various sizes.

However, it should be stressed that poor prediction performance also occurs in more classical MLR PTFs which use only texture as predictor variable. This is particularly the case if the training dataset was vastly different from the test dataset, e.g., in terms of OM content and type which can play a role on soil wettability (e.g., Matthews et al., 2008). In various PTFs, soil OC content has indirectly accounted for soil wettability as indicated by Hwang et al. (2011) when referring to the study conducted by Rawls et al. (2003).

However, it is good to mention that the prediction performance of the symmetric PSF model of Hwang et al. (2011) is by far better than the one of many temperate and some tropical climate MLR PTFs when applied on ‘tropical’ soils. This was confirmed in Chapter 5 dealing with prediction performance of tropical and temperate MLR water retention PTFs on Lower Congo soils.

Table 9-5. Performance in terms of MD, RMSD and R^2 for the symmetric PSF model of Hwang et al. (2011) to predict water retention of Lower Congo soils (test dataset) based on the IGBP-Trop-2 dataset (calibration dataset).[†]

	Matric potential							
	-1 kPa	-3 kPa	-6 kPa	-10 kPa	-20 kPa	-33 kPa	-100 kPa	-1500 kPa
Coarse textured								
MD	0.005	-0.055	0.000	0.016	0.033	0.022	-0.005	-0.033
RMSD	0.054	0.083	0.053	0.059	0.061	0.057	0.049	0.055
R^2	0.203	0.035	0.391	0.388	0.342	0.279	0.291	0.205
Medium textured								
MD	0.012	-0.018	-0.023	-0.030	-0.029	-0.047	-0.053	-0.065
RMSD	0.046	0.064	0.076	0.078	0.066	0.076	0.074	0.083
R^2	0.690	0.463	0.344	0.353	0.486	0.483	0.549	0.374
Fine textured								
MD	0.012	0.012	0.019	0.020	0.025	0.016	0.011	0.003
RMSD	0.035	0.056	0.071	0.078	0.084	0.087	0.085	0.085
R^2	0.648	0.304	0.130	0.082	0.108	0.07	0.153	0.297

[†] MD (in $\text{m}^3 \text{m}^{-3}$) is mean difference; RMSD (in $\text{m}^3 \text{m}^{-3}$) is root mean square difference; R^2 (-) is coefficient of determination.

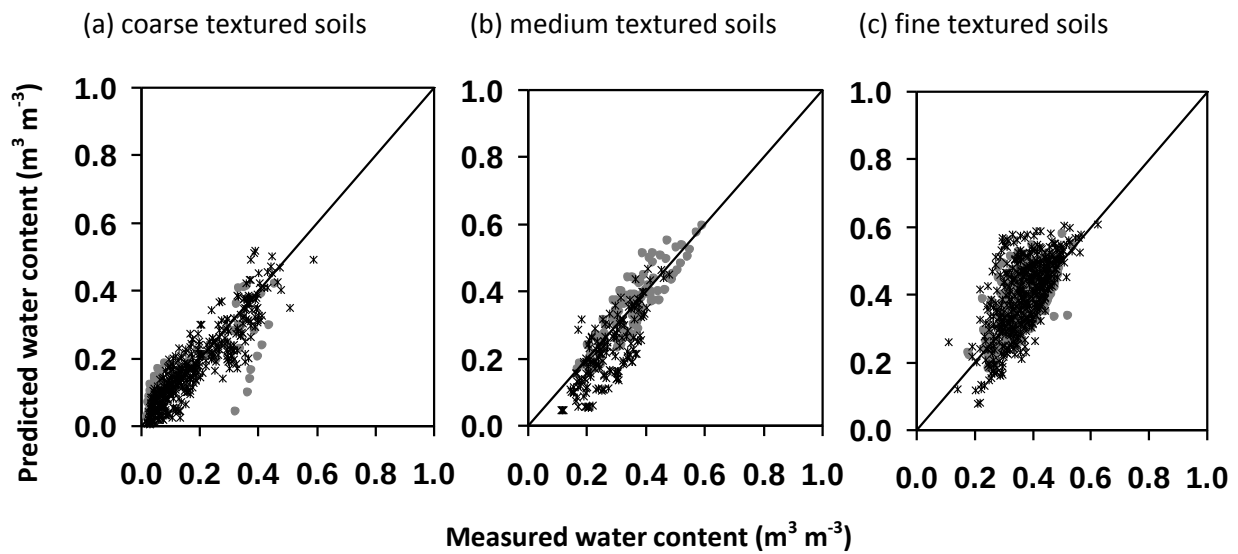


Figure 9–2. Measured versus predicted soil water contents by the symmetric pore-solid fractal model of Hwang et al. (2011) for 196 highly weathered soils from the Lower Congo dataset used as validation dataset. Grey circles represent unimodal soils and black crosses represent bimodal soils.

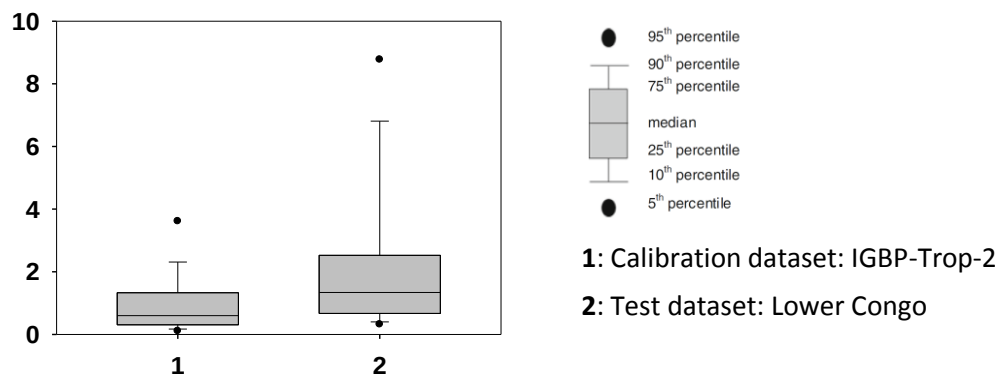


Figure 9–3. Box plots of organic carbon content (%) of (1) IGBP-Trop-2 (calibration dataset, 209 soil samples) and (2) Lower Congo (test dataset, 196 soil samples).

9.3.4 Prediction performance of the symmetric PSF model tested on Lower Congo soils using k-fold cross validation

From the previous test, limitations were observed in the reliability, i.e., the prediction performance of the symmetric PSF model of Hwang et al. (2011) when calibrated on an independent dataset of highly weathered soils. We therefore performed a k-fold cross-validation (Pachepsky and Schaap, 2004) using the Lower Congo dataset only. This was done in order to better appreciate the accuracy, i.e., the predictive capacity of the symmetric PSF model when a portion of the Lower Congo dataset is used as calibration dataset and the remaining portion as validation dataset. The predictive performance of the symmetric PSF model from the cross-validation procedure (accuracy) was similar ($0.037 \text{ m}^3 \text{m}^{-3} < \text{average RMSD} < 0.080 \text{ m}^3 \text{m}^{-3}$) compared to the predictive performance (reliability) ($0.035 \text{ m}^3 \text{m}^{-3} < \text{RMSD} < 0.087 \text{ m}^3 \text{m}^{-3}$) with the independent calibration procedure as can be seen in Table 9–5 and

Table 9–6. This means that the predictive performance of the symmetric PSF model is strongly dependent on the similarity between the calibration and the validation dataset.

The results yielded by the cross-validation procedure showed that the prediction performance of the symmetric PSF model was particularly poor for two or three k-fold validation subsets out of 10 or 11 with RMSD values greater than $0.100 \text{ m}^3 \text{ m}^{-3}$ in each aggregated FAO textural group. This has a marked influence on the overall average RMSD values at different matric potentials calculated across 10-11 folds particularly for fine-textured soils. This was further put in evidence by the difference between the average and the median RMSD values at various matric potentials. This shows that these two or three subsets were vastly different from the others in e.g., their structural properties (the geometry and distribution of the pore system) and subsequently in their packing-related parameter γ and in their OC content as well. The prediction performance of the PSF model based on internal cross-validation seems to decrease from coarse to fine textured soils.

As indicated previously, the complexity of the porous system especially in fine textured soils can explain the low prediction performance of semi-physical models which consider exclusively (detailed) texture as sole predictor of water retention. Rawls et al. (2003) investigated the effect of soil OC on soil water retention based on a subset of about 12,000 samples extracted from the National Soil Characterization database (Soil Survey Staff, 1995). They found that the effect of changes in OC content on soil water retention depends on the proportion of textural components and the amount of OC present in the soil. Water retention of soils with coarse texture is substantially more sensitive to the amount of OC as compared with fine-textured soils. However, overall median RMSD values after cross-validation were acceptable with values $< 0.070 \text{ m}^3 \text{ m}^{-3}$.

We suggest that in future studies, the packing-related parameter γ should be optimized for the five aggregated USDA textural class and even for the 12 USDA textural classes for soils belonging to the same WRB Reference Soil Group. A larger database should be used to have a representative number of soil samples per textural class considered. Moreover, the influence of the OM content on the optimized values of the packing-related parameter γ for each textural class can be also investigated. This may help to improve the predictive performance of the symmetric PSF model for soils of the humid tropics. Another point of future research could be the consideration of more realistic wettability properties and adsorptive forces generated by OM, kaolinite, Fe and Al content of highly weathered soils in semi-physical models.

Table 9-6. Performance in terms of RMSD values for the symmetric PSF model of Hwang et al. (2011) to predict water retention of Lower Congo soils based on a k-fold cross-validation.†

Matric potential								
	−1 kPa	−3 kPa	−6 kPa	−10 kPa	−20 kPa	−33 kPa	−100 kPa	−1500 kPa
	RMSD							
	Coarse textured soils							
Fold 1	0.049	0.118	0.101	0.111	0.100	0.110	0.126	0.137
Fold 2	0.030	0.067	0.029	0.043	0.052	0.033	0.010	0.029
Fold 3	0.024	0.068	0.046	0.041	0.029	0.022	0.034	0.057
Fold 4	0.028	0.049	0.042	0.034	0.043	0.030	0.008	0.030
Fold 5	0.078	0.063	0.072	0.069	0.070	0.052	0.028	0.033
Fold 6	0.077	0.032	0.032	0.027	0.030	0.023	0.029	0.058
Fold 7	0.107	0.079	0.075	0.063	0.061	0.055	0.030	0.035
Fold 8	0.045	0.095	0.037	0.042	0.037	0.033	0.031	0.047
Fold 9	0.052	0.118	0.024	0.053	0.045	0.033	0.012	0.012
Fold 10	0.052	0.082	0.035	0.075	0.085	0.083	0.036	0.009
Fold 11	0.061	0.054	0.056	0.071	0.099	0.092	0.061	0.014
Average RMSD	0.055	0.075	0.050	0.057	0.059	0.051	0.037	0.042
Median RMSD	0.052	0.068	0.042	0.053	0.052	0.033	0.030	0.033
	Medium textured soils							
Fold 1	0.020	0.019	0.083	0.072	0.030	0.012	0.026	0.059
Fold 2	0.041	0.041	0.038	0.030	0.037	0.032	0.043	0.085
Fold 3	0.047	0.041	0.042	0.042	0.034	0.035	0.045	0.069
Fold 4	0.100	0.080	0.065	0.057	0.037	0.043	0.025	0.006
Fold 5	0.051	0.056	0.061	0.063	0.027	0.035	0.038	0.015
Fold 6	0.026	0.028	0.042	0.051	0.033	0.057	0.051	0.038
Fold 7	0.039	0.045	0.052	0.058	0.068	0.074	0.066	0.060
Fold 8	0.032	0.087	0.108	0.118	0.097	0.110	0.110	0.128
Fold 9	0.059	0.118	0.140	0.148	0.145	0.164	0.151	0.153
Fold 10	0.035	0.076	0.089	0.095	0.081	0.096	0.093	0.099
Fold 11	0.010	0.059	0.056	0.047	0.039	0.047	0.051	0.069
Average RMSD	0.042	0.059	0.070	0.071	0.057	0.064	0.064	0.071
Median RMSD	0.039	0.056	0.061	0.058	0.037	0.047	0.051	0.069
	Fine textured soils							
Fold 1	0.039	0.049	0.050	0.058	0.073	0.062	0.075	0.094
Fold 2	0.046	0.101	0.140	0.156	0.157	0.157	0.161	0.161
Fold 3	0.039	0.066	0.090	0.100	0.119	0.122	0.116	0.112
Fold 4	0.046	0.070	0.090	0.101	0.113	0.118	0.111	0.105
Fold 5	0.020	0.031	0.039	0.041	0.044	0.042	0.047	0.046
Fold 6	0.045	0.069	0.077	0.080	0.079	0.076	0.062	0.040
Fold 7	0.064	0.058	0.058	0.058	0.061	0.051	0.043	0.036
Fold 8	0.031	0.035	0.037	0.039	0.031	0.059	0.046	0.037
Fold 9	0.029	0.022	0.035	0.042	0.036	0.037	0.042	0.047
Fold 10	0.026	0.026	0.034	0.040	0.052	0.076	0.061	0.071
Average RMSD	0.039	0.053	0.065	0.071	0.077	0.080	0.077	0.075
Median RMSD	0.039	0.053	0.054	0.058	0.067	0.069	0.062	0.059

† RMSD is root mean square difference ($\text{m}^3 \text{m}^{-3}$).

9.4 Conclusion

This preliminary study investigated the applicability of three PSF models to describe and predict water retention phenomena for highly weathered soils of the humid tropics. A strong relationship was found between texture (sand and clay content) and the fractal dimensions of PSD curve and SWRC for soils of the humid tropics. Based on a dataset of 196 soils from Lower Congo, the symmetry-based PSF model of Hwang et al. (2011) was found to better describe water retention of highly weathered soils compared to the asymmetry-based PSF model and the original PSF model of Bird et al. (2000). Therefore, the assumption of symmetry between the PSD and POD seems to hold in the case of highly weathered soils. This was contrary to the results found by Hwang et al. (2011) on temperate-climate soils. However, further investigations with soils from other tropical regions need to be conducted to confirm the results of this study.

On the other hand, a calibration procedure was performed based on an independent dataset of highly weathered soils to investigate the potential of the symmetric PSF model of Hwang et al. (2011) to predict water retention at various matric potentials. A limited prediction performance was found but this may be attributed to differences in pore geometry and in OM content between the training dataset and the independent test dataset. Applying a k-fold cross-validation procedure using only Lower Congo soil data, a better prediction performance of the symmetric PSF model could be reached. These results showed the importance of a better calibration of the packing-related parameter γ to improve the prediction performance of the PSF model. It is important to note that the results found in this study are non conclusive. More investigations need to be conducted in order to better evaluate and/or improve the potential of the PSF theory to explain water retention phenomena for highly weathered soils. Improvement may be possible by the application of a more realistic model for pore space geometry and by considering the individual contribution of adsorptive and capillary forces to the matric potential. So far, studies related to the PSF models were limited to temperate-climate soils. The marked differences in physico-chemical, structural, mineralogical and hydraulic properties between ‘tropical’ and ‘temperate’ soils need to be further investigated in the light of the PSF theory.

10 Evaluation of pedotransfer functions for mapping soil hydrophysical properties in Lower Congo

10.1 Introduction

Soil parameters that represent hydraulic properties of different soils are essential for a wide range of modelling studies, e.g. such as crop growth and yield prediction, evaluation of agroforestry systems, and models describing relationships between soil, vegetation and atmosphere. The latter are vital components of General Circulation Models, describing the land surface–atmosphere interactions (Hodnett and Tomasella, 2002), and require (soil) hydraulic information on a large spatial scale (Tomasella and Hodnett, 2004).

Soil bulk density (BD) is defined as the mass of an oven-dry sample of undisturbed soil per unit bulk volume (ISSS Working Group, 1998, p.153). As discussed in Chapter 7, bulk density (BD) is a key soil property for sustainable land management in various agro-climatic zones in the world. The influence of many soil management practices on soil properties that affect water movement and water retention in soils can be reflected in the soil BD (Kaur et al., 2002). Furthermore, soil organic carbon (OC) estimates require data on BD for weight-to volume conversion (e.g., for OC, nutrients, and trace elements) and for assessing stocks and fluxes (Martin et al., 2007). The soil carbon stock C_t (g cm^{-2}) is calculated as follows (Howard et al., 1995; Arrouays et al., 2001; Benites et al., 2007):

$$C_t = BD (\text{g cm}^{-3}) \times OC \text{ concentration } (\%) \times \text{sampling depth (cm)} \quad [10-1]$$

For instance, models dealing with water flow or solute transport are in general very sensitive to bulk density (Brown and Heuvelink, 2005). Recently, Hollis et al. (2012) wrote that BD is a critical soil characteristic for soil studies and modeling and assessment of soil carbon and nutrient stocks or environmental risks is required for implementation of various environmental policies in Europe.

The soil water holding capacity (SWHC) is defined as the amount of water retained between field capacity (FC) and permanent wilting point (PWP). Estimated in undisturbed soil cores, FC is approximated by the water content determined at -10 or -33 kPa depending on physical conditions. On the other hand, the PWP is usually referred to as the water content at -1500 kPa. The SWHC is a key indicator which is widely used to determine the suitability of a given region for producing a given crop (Sys et al., 1991, Givi et al. 2004), or to assess the environmental risks within a region (Cazemier et al., 2001). It is an essential component of the water and energy balances of the terrestrial biosphere and represents one of the most important soil factors for plant growth, influencing carbon allocation, nutrient cycling, and the rate of photosynthesis (Milly and Shmakin, 2002; Piedallu et al., 2011; Hong et al., 2013). Characterisation of SWHC is necessary for assessing the soil's physical status and quality. With environmental issues related to climate change and agriculture, the SWHC is a key element that should be taken into account to increase water use efficiency in view of the expected decrease in rainfall and soil water availability in various regions of the world (FAO, 2002).

However, there are few studies that have mapped soil hydrophysical properties, i.e., BD, soil water retention curve (SWRC), and hydraulic conductivity (K) as indicated by Hong et al. (2013). This is particularly true for countries, such as the D.R. Congo, that lack data on hydrophysical soil properties in their soil survey reports and databases. Measurement of these hydrophysical characteristics (BD, SWRC and K) governing retention and transport of water and chemicals in soils, is known to be labour-intensive, time-consuming and expensive, particularly for large scale studies. The paucity of reliable information on soil hydraulic parameters has been identified as a major constraint in the application of hydrological modelling techniques at a regional scale (D'Urso and Basile, 1997; Wösten et al., 2013).

Thus, there is a need to resort to estimating hydraulic characteristics from other more readily available soil data through the application of pedotransfer functions (PTFs) (Bouma, 1989; Wösten et al., 2001). Few studies present PTFs developed for ‘tropical’ soils compared to ‘temperate’ soils as indicated in Chapter 3 and by various researchers (Nebel et al., 2010; Minasny and Hartemink, 2011; Wösten et al., 2013). To the limit of our knowledge, no hydrophysical PTFs were developed for soils located in Central Africa. In Chapter 5, the predictive ability of some ‘temperate’ and ‘tropical’ PTFs was evaluated based on a limited dataset of Lower Congo soils and ‘temperate’ PTFs were found to overestimate water content at -33 kPa and -1500 kPa. In Chapters 6, 7, 8 and 9, several approaches of pedotransfer modelling were tested to predict BD and water retention at various matric potentials for soils of Lower Congo. In the preceding chapters, comparisons between various PTFs and approaches of pedotransfer modelling were made based on predicted and measured BD and water retention values. This chapter investigates the practical use of PTF-estimated values to describe the spatial variability of hydrophysical properties of Lower Congo soils.

Soil water management models are used in various agricultural and environmental studies to investigate and suggest different scenarios for a better management of natural resources. Simulation models, when used at regional scale, have to be up-scaled from the point soil profiles (Hopmans, 1987). This can be done using geostatistical methods (Van Meirvenne et al., 1995), or effective hydraulic parameters (Smith and Diekkrüger, 1996). The drawback of the latter option is that it does not reflect the spatial pattern of the soil water balance elements because it interprets the whole field as an equivalent soil profile (Farkas et al., 2008). Van Meirvenne et al. (1994) performed the first study based on geostatistical analyses to provide a continuous map of sand content of the Lower Congo region (designated as Bas-Zaïre). They quantified soil textural fractions by comparing three methods to predict clay and sand content using a reconnaissance soil map at 1:1,000,000 scale and 151 sampled soil profiles. They provided a quantitative map of the sand content of Bas-Congo using a combination of choropleth maps, point observations and kriging interpolations.

Romano and Santini (1997) evaluated some published PTFs related to their ability to quantify the spatial structure and variability of soil water retention for some ‘temperate’ soils in Italy. Based on geostatistical analyses, they concluded that there are strong similarities in the structure of the spatial variability detected by the PTFs. However, they indicated that the above conclusions should be confirmed by similar experiments in other test sites. Nebel et al. (2010) conducted a similar study for lowland soils of Brazil by assessing the predictive ability of eight point-based PTFs to predict water content at -33 kPa and -1500 kPa and their potential in describing the structure of spatial variability of these hydraulic characteristics.

Due to the large size of the study area, the limited number of point soil profiles, the irregular distribution of the sampling scheme and the lack of measured soil hydrophysical properties, the focus of this study is on evaluation of PTFs rather than on the production of high-quality soil maps. Therefore, this chapter describes the influence of various PTFs, approaches to pedotransfer modelling, and spatial interpolation techniques on the level of uncertainty when producing continuous maps of BD and SWHC for Lower Congo. The objective of this study is to perform a functional evaluation of selected PTFs in spatially describing hydrophysical properties of Lower Congo soils.

10.2 Materials and Methods

10.2.1 Study area: Lower Congo

The study area has been described in Chapter 2.

10.2.2 Datasets

Two datasets based on soil samples taken by horizon up to 100 cm depth have been included in the analysis: (1) a recent survey of 24 soil profiles, sampled within the frame of this study, and (2) a legacy dataset of 248 soil profiles.

10.2.2.1 Recently surveyed soil profiles

In the autumn of 2009, 155 undisturbed soil samples were collected from 24 soil profiles located in Lower Congo (Figure 10–1). The database comprises mainly highly weathered soils as well as coarse textured soils that developed under humid tropical conditions and are classified as Ferralsols, Acrisols, Nitisols and Arenosols (IUSS Working Group WRB, 2006). Soil samples were collected under different land uses (savannah, forest, arable land and old quarries) and they cover a large range of textures (see Chapters 6 to 9). Bulk density was determined on undisturbed soil samples and soil water content was determined from the wet to the dry range at eight different matric potentials: -1 , -3 , -6 , -10 , -20 , -33 , -100 and -1500 kPa. More details on the lab analyses are given in Chapters 4 to 9. This dataset was used both as training and test (validation) dataset to develop BD and water retention PTFs for Lower Congo soils.

10.2.2.2 Legacy soil data

This dataset is derived from 248 soil profiles which were collected during soil survey campaigns which took place after re-launching the systematic soil survey of Lower Congo in the 1980s. The 972 soil samples considered in this study have been taken within 100 cm depth. Data on particle size distribution (PSD), OC and pH are available in digital format for 972 soil samples whereas data on CEC, DCB-Fe and DCB-Al are only available in digital format for 385 soil samples (99 profiles). There are no data on BD, SWRC or K in this dataset.

The spatial distribution of the 24 profiles with measured BD and SWRC and the 248 profiles without hydrophysical data are presented in Figure 10–1. It can be seen that the soil profiles are highly clustered and irregularly distributed over the entire Lower Congo region. It should be noted that in this study, SWRC was not adjusted for the gravel content as data on mass fraction of gravel (kg kg^{-1}) were not available in the Lower Congo dataset. Soil samples were not taken at the same depth for physico-chemical and hydrophysical analyses. The soil sampling scheme in Lower Congo was not spatially uniform in all the three spatial dimensions as it was a direct product of somewhat independent survey campaigns. In Lower Congo, soil sampling campaigns were dictated by various objectives: determination of soil nutrient status, mapping, mineralogical analyses, land evaluation, pedotransfer modelling among other things. Van Meirvenne et al. (1994) conducted the first geostatistical study to quantify soil textural fractions and compared three methods to predict clay and sand content in the Bas-Congo region (formerly known as Bas-Zaïre) using a reconnaissance soil map and 151 soil profiles dug during a free survey of the region.

A simple stratification was performed dividing the Lower Congo region into two main zones which are largely related to the geology of the region: the ‘clayey’ area and the ‘sandy’ area (Figure 10–1). The ‘clayey’ area which is the largest zone corresponds more or less to the axial region described by Stoops (1966). The ‘sandy’ area covers two small regions: the coastal and the oriental regions separated by the axial region (Stoops, 1966). The ‘clayey’ region dominated by fine textured soils has 204 and 16 profiles from the previous and recent surveys, respectively, whereas the ‘sandy’ region, dominated by coarse textured soils, has 44 and 8 profiles, respectively.

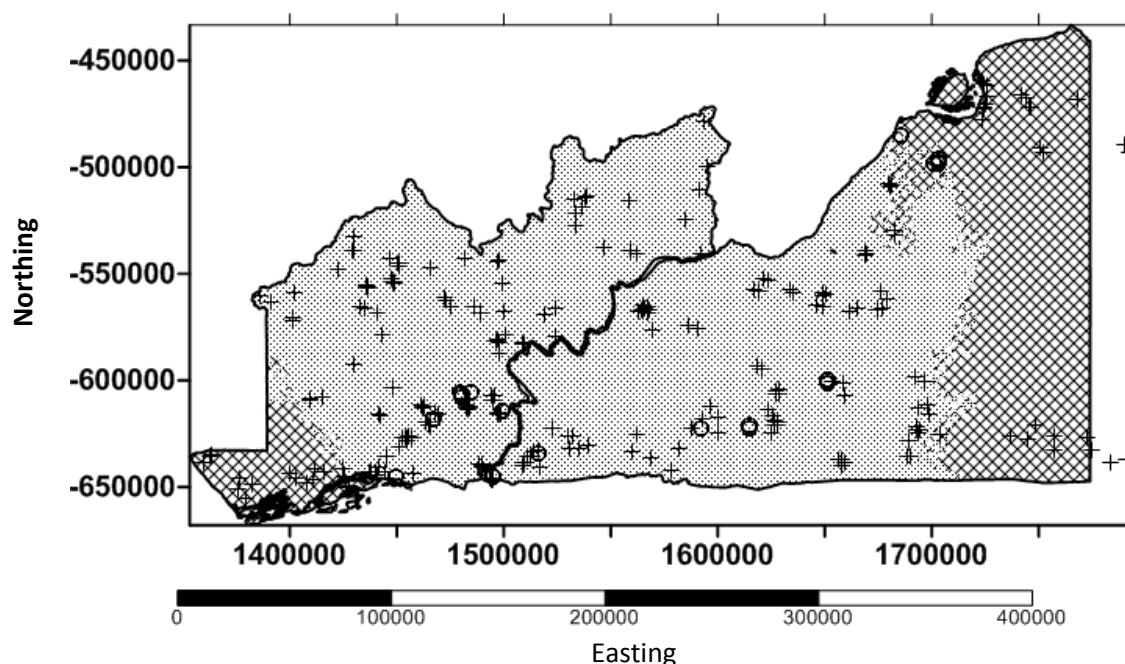


Figure 10–1. Location of the soil profiles and stratification of Lower Congo in two main textural zones. Crosses represent the locations of the 248 soil profiles issued from previous survey and for which only primary soil properties were available, whereas circular points represent the locations of the 24 soil profiles issued from this study and for which also hydrophysical properties have been determined. The two regions in diagonal cross represent the ‘sandy’ stratum, whereas the central region in fine points represents the ‘clayey’ stratum.

10.2.3 Site-specific pedotransfer modelling approaches

10.2.3.1 Soil bulk density

Pedotransfer functions developed in Chapter 7 for Lower Congo soils were used to estimate BD for the 248 locations. Two different approaches of pedotransfer modelling were tested: (1) an equation-based approach where multiple linear regression (MLR) equations are developed using a stepwise regression procedure, (2) and an approach based on pattern-recognition, called the k-nearest neighbour (k-NN) method. Due to limitation in available soil information on CEC, DCB-Fe and DCB-Al, only texture and OC were considered as predictors in both approaches. Details on PTFs for BD using MLR and k-NN methods are given in Chapter 7 and in Table 10–1.

10.2.3.2 Soil water retention

Various approaches for predicting water content at -10 and -1500 kPa in order to estimate the SWHC were selected in this study. In Chapters 8 and 9, site-specific PTFs for Lower Congo soils were developed based on three different approaches: (1) an equation-based approach with MLR, (2) an approach based on pattern-recognition with k-NN and (3) an approach based on the pore-solid fractal (PSF) theory. Some of these equation-based PTFs which use predictors such as texture (SSC), OC and pH have been selected for the purposes of this study and are reported in Table 10–1. As indicated in Chapter 9, the symmetric PSF model of Hwang et al. (2011) requires both detailed PSD and BD to convert PSD into pore size distribution (POD). To be able to test the PSF approach, three average BD values calculated based on the Lower Congo dataset of 196 soil samples (see Chapter 7) were considered: 1.37 g cm^{-3} , 1.34 g m^{-3} and 1.46 g cm^{-3} for the FAO-UNESCO (1974) fine, medium and coarse textured soils, respectively.

10.2.4 Published pedotransfer functions for water retention

Additionally, some published PTFs were selected for comparative purposes based on the availability of soil information that can be used as potential predictors. As previously mentioned, the most available soil properties in Lower Congo datasets are sand, silt, clay, OC and pH. Due to limitation of soil information such as BD and CEC that could be used as inputs, other well-known ‘tropical’ PTFs such as parameter-based PTFs of Hodnett and Tomasella (2002) or point-based PTFs of Minasny and Hartemink (2011) could not be tested in this study. Eventually, the ‘tropical’ PTFs of van den Berg et al. (1997) and the ‘temperate’ PTFs of Rawls and Brakensiek (1982) tested in Chapter 5 were considered. These point-based PTFs were selected as they require only information on clay, sand and OC content to predict water content at –10 and –1500 kPa (Table 10–1).

Table 10-1. Site-specific and published multiple linear regression pedotransfer functions used in this study

Pedotransfer functions	Geographical origin
Bulk Density (Mg m^{-3})	
Chapter 7 $BD = 1.64581 - 0.00362 \times \text{Clay} - 0.0016 \times \text{Sand} - 0.0158 \times \text{OC}$	Lower Congo
Water content at –10 kPa (vol%)	
Chapter 8 $\theta_{-10\text{kPa}} = 43.520 - 0.296 \times \text{Sand}$ $\theta_{-10\text{kPa}} = 55.016 - 0.289 \times \text{Sand} - 2.275 \times \text{pH}$	Lower Congo
van den Berg et al. (1997) $\theta_{-10\text{kPa}} = 10.880 + 0.347 \times \text{Clay} + 0.211 \times \text{Silt} + 1.756 \times \text{OC}$	humid tropics of various countries
Rawls and Brakensiek (1982) $\theta_{-10\text{kPa}} = 41.180 - 0.300 \times \text{Sand} + 0.230 \times \text{Clay} + 3.170 \times \text{OM}$	temperate regions in US
Water content at –1500 kPa (vol%)	
Chapter 8 $\theta_{-1500\text{kPa}} = 21.046 - 0.170 \times \text{Sand} + 0.175 \times \text{Clay}$ $\theta_{-1500\text{kPa}} = 28.004 - 0.156 \times \text{Sand} + 0.188 \times \text{Clay} - 1.564 \times \text{pH}$	Lower Congo
van den Berg et al. (1997) $\theta_{-1500\text{kPa}} = 3.830 + 0.272 \times \text{Clay} + 0.212 \times \text{Silt}$	humid tropics of various countries
Rawls and Brakensiek (1982) $\theta_{-1500\text{kPa}} = 2.600 + 0.500 \times \text{Clay} + 1.580 \times \text{OM}^\dagger$	temperate regions in US

† OM = organic matter = 1.724 OC

10.2.5 Calculation of depth-weighted averages of clay, sand, OC, BD, and SWHC

Depth-weighted averages of clay, sand and OC contents within 100 cm depth (designated here as Clay_{avg} , Sand_{avg} and OC_{avg} , respectively) were computed for each of the 24 and 248 Lower Congo soil profiles. A depth-weighted average BD within 100 cm depth (designated here as BD_{avg}) was computed for each of the 24 profiles with lab-determined BD values. For each of 248 profiles with unknown BD values, site-specific PTFs developed in Chapter 7 were applied to first derive BD at each layer and

subsequently the corresponding estimated depth-weighted average was computed. Due to the relative homogeneity of highly weathered soils and the relatively low OC content in the topsoil, no distinction was made between the top and subsoil layers.

In this study, water content at -10 kPa was considered as representing field capacity (FC) and water content at -1500 kPa as representing the permanent wilting point (PWP). As explained in Chapter 3, various researchers (Pidgeon, 1972; Babalola, 1979; Lal, 1978; Reichardt, 1988), considered the water content at -33 kPa to be too low to represent FC for ‘tropical’ soils and they suggested the measure of soil moisture at -10 kPa.

The profile soil water holding capacity (designated here as PSWHC, in cm/m depth) is calculated as the sum of SWHC in all the layers present in a soil profile within 100 cm depth as follows:

$$PSWHC = \sum_{i=1}^{N_{lyr}} SWHC_i = \sum_{i=1}^{N_{lyr}} (\theta_{-10kPa} - \theta_{-1500kPa})_i \times d_i \quad [10-2]$$

with $\theta_{x\text{ kPa}}$ = volumetric water content at x kPa in $\text{m}^3 \text{m}^{-3}$

N_{lyr} = number of layers within 100 cm depth.

d_i = thickness of the i^{th} layer (cm)

Similarly to BD_{avg} , PSWHC was calculated for the 24 profiles with lab-determined water retention values. For each of the 248 profiles with unknown water retention values, selected PTFs were applied to estimate water content at -10 and -1500 kPa and to derive the corresponding PSWHC values.

10.2.6 Use of existing Lower Congo soil maps

Previously existing digitized maps of Lower Congo (Van Ranst et al., 2010) available at the Laboratories of Soil Science at the Ghent University (Belgium) and at the University of Kinshasa (D.R. Congo) were used as base or thematic maps to provide a correct geographical location for all the soil profiles and a textural stratification of the area. All the existing maps were processed in ArcGIS 10.0 (Environmental Systems Research Institute, Redlands, CA). The existing thematic maps used in this study were at scale 1:500,000 and were projected using a Mercator projection secant to the parallels 5° N and 5° S.

10.2.7 Mapping of physico-chemical and hydrophysical properties

10.2.7.1 Preliminary statistical exploration of the target soil properties

The first step was to assess normality of the data to be mapped. This analysis was performed using a Kolmogorov-Smirnov test and visually using histograms based on data of the 248 soil profiles. As previously mentioned, the physico-chemical properties as $Sand_{avg}$, $Clay_{avg}$ and OC_{avg} content were measured (legacy data), whereas the hydrophysical properties (BD_{avg} and PSWHC) were secondary data (derived from legacy data through PTFs).

10.2.7.2 Spatial interpolation methods

In their review study on spatial interpolation methods in environmental sciences, Li and Heap (2011) investigated the performance and impact factors of popular interpolation methods in environmental sciences and concluded that inverse distance weighting (IDW) and ordinary kriging (OK) were the most popular ones. In order to get predictions of BD and PSWHC at every location within the study area using 248 estimated BD and PSWHC points, these two most used interpolation methods were applied: IDW and OK. The BD_{avg} and PSWHC values estimated by the aforementioned PTFs were interpolated over the whole area of Lower Congo.

For the IDW and OK interpolation methods, the value of a given variable Z at the unsampled location x_0 , $Z^*(x_0)$ is estimated based on the data from the surrounding locations, $Z(x_i)$, as:

$$Z^*(x_0) = \sum_{i=1}^{N_b} w_i Z(x_i) \quad [10-3]$$

where w_i is the weight assigned to each $Z(x_i)$ value and N_b is the number of the closest neighbouring sampled data points used for estimation. The weights for IDW are usually proportional to the inverse of the squared distance between the prediction point and the observation points, and they sum to 1, as follows:

$$w_i = \frac{d_i^{-2}}{\sum_{i=1}^{N_b} d_i^{-2}} \quad [10-4]$$

where d_i is the distance between the estimated point and the observed point.

On the other hand, OK calculates the values of w_i by estimating the spatial structure of the variable's distribution represented by a sample semi-variogram as:

$$\gamma(l) = \frac{1}{2N_p(l)} \sum_{i=1}^{N_p(l)} [Z(x_i) - Z(x_i + l)]^2 \quad [10-5]$$

where $N_p(l)$ is the number of data pairs separated by the distance l , x_i and x_{i+1} are sampling locations separated by a distance l , and $Z(x_i)$ and $Z(x_i + l)$ are the observed values of variable Z at the corresponding locations.

The semi-variance $\gamma(l)$ is equal to half the expected squared difference between values at locations separated by a given lag l and used to express spatial variation (Journel and Huijbregts, 1987). The semi-variance was calculated for various l values and data points of the experimental semi-variogram were fitted with a theoretical model and the most appropriate models for the purpose of this study were explored. Depending on the observed or PTF-estimated soil property, isotropic and omnidirectional linear, spherical or exponential models were selected.

The linear semi-variogram corresponds to the following formula:

$$\gamma^*(l) = C_o + C l; \text{ if } l \geq 0 \quad [10-6]$$

$$\gamma^*(l) = 0; \text{ otherwise}$$

The spherical semi-variogram is expressed as follows:

$$\gamma^*(l) = C_o + C \left[\frac{1.5l}{a} - 0.5 \left(\frac{l}{a} \right)^3 \right]; \text{ if } 0 \leq l \leq a \quad [10-7]$$

$$\gamma^*(l) = C_o + C; \text{ if } l > a,$$

$$\gamma^*(l) = 0; \text{ otherwise}$$

The exponential semi-variogram corresponds to the following formula:

$$\gamma^*(l) = C_o + C \left(1 - \exp \left(-\frac{l}{a} \right) \right); \text{ if } l \geq 0 \quad [10-8]$$

$$\gamma^*(l) = 0; \text{ otherwise}$$

where a is the range, C_0 is the nugget, C_0+C is the sill and l is the distance separating two observations.

Each semi-variogram model is thus described by three main parameters: range of spatial correlation, nugget, and sill. The range of the spatial correlation measures the distance beyond which two observations are spatially uncorrelated. The nugget represents measurement error and unobserved microscale variability. The sill represents the constant variance of the observations for second-order stationary spatial processes. More details on these semi-variogram parameters can be found in Isaaks and Srivastava (1989), Cressie (1993), and Goovaerts (1997).

When comparing IDW and OK for continuously mapping PSWHC, a unique theoretical model (spherical) was selected to allow an easy and direct comparison of the effects of different prediction approaches, PTF models and set of predictors on the quality of the interpolation using cross-validation. In previous studies, dealing with the ability of various PTFs to describe the spatial distribution of soil hydraulic properties, a spherical model was found to be the most suitable to apply OK (Romano and Santini, 1997; Ferrer-Julià et al., 2004; Farkas et al., 2008; Nebel et al., 2010). In the case of IDW, no semi-variogram or theoretical model is required. To assess the predictive quality of the two interpolation procedures, two statistical indices were used: root mean squared difference (RMSD) as an expression of the overall error and coefficient of determination (R^2) as an expression of goodness-of-fit.

Based on the dimensions of the study area, a maximum lag of 80,000 m was retained for all the semi-variograms, similar to the previous work of Van Meirvenne et al. (1994) of the Bas-Congo region. For all these analyses, a spatial resolution of 1 km x 1 km was considered in order to take the maximum benefit from soil data. This cell size was selected after computing the number of soil profiles per square kilometer of the study area and taking into account the density of the profile network. Though the average is 1 soil profile every 235 km², there are several areas where the profiles are within a distance of less than 1 km, particularly along catenas or transects. The same 1 km x 1 km resolution was taken by Ferrer-Julià et al. (2004) when constructing a saturated hydraulic conductivity map of Spain using PTFs and spatial prediction.

10.2.7.3 Performance and validation of the interpolated maps

The performance of the interpolation procedure was evaluated through cross-validation following a leave-one-out procedure. The latter consists in removing one point from the dataset, and using the remaining data and the specified method (IDW or OK) to interpolate a value at that point. Then the process is repeated for the next point.

To assess the predictive quality of the final maps, an independent validation was performed. The observed (calculated) BD_{avg} and PSWHC values of the 24 profile locations that were not included in the interpolation process were compared with the values yielded by the interpolation of the estimated PSWHC values at the 248 locations.

Mean difference (MD), root mean square difference (RMSD) and the coefficient of determination (R^2) were used to assess the prediction performance of the interpolation methods and the influence of the different PTFs on the uncertainty of the derived maps. The aforementioned statistical indices and their related formulas have been introduced in previous chapters.

After selecting the best interpolation method along with the most suitable PTFs, continuous maps of BD_{avg} and PSWHC, respectively, were constructed including the 24 sample locations with observed values of BD_{avg} and PSWHC.

Surfer 10 (Golden Software, Inc., Golden, CO) was used to conduct statistical exploration, do IDW and OK interpolations, evaluate the predictive ability of each interpolation procedure, perform an independent validation and construct the final maps of BD_{avg} and PSWHC for Lower Congo.

10.3 Results and discussion

10.3.1 Preliminary statistical exploration of the target soil properties

Application of OK is based upon the presumption that the data to be interpolated should be normally distributed. Therefore a quick statistical exploration was performed on the primary data i.e. $Clay_{avg}$, $Sand_{avg}$ and OC_{avg} using histograms. The textural properties $Clay_{avg}$ and $Sand_{avg}$ did not depart much from normality. On the other hand, $Silt_{avg}$, pH_{avg} , OC_{avg} values showed a skewed distribution suggesting a log-transformation of the data to achieve a more normal distribution as described in Figure 10–2 for OC_{avg} . Mean, median, standard deviations, minimum and maximum values and coefficients of variation (CV) for $Clay_{avg}$, $Sand_{avg}$ and OC_{avg} for the 248 and the 24 profiles are shown in Table 10–2. For BD_{avg} and PSWHC, only statistics of the 24 profiles with ‘observed’ values are reported. Regarding the PTF-estimated values of BD_{avg} and PSWHC for each of the 248 profiles, their descriptive statistics differ from one PTF to the other.

Variation in textural composition observed in this study is in agreement with a previous study conducted by Van Meirvenne et al. (1994). They used a subset of 151 profiles taken from the 248 profiles used in this study and observed extreme differences in textural composition. Among all soil properties, OC_{avg} exhibited the highest variation (CV = 166%). This largely explains the non-normality of its distribution and the necessity to perform a log-transformation of the data (Figure 10–2). On the other hand, a substantial variation of PSWHC values was also observed (CV = 33.9%) for the 24 profiles compared to BD_{avg} (CV = 9.2%) which is not surprising considering the ranges of variation of these soil properties.

For sake of simplicity, only the 248 profiles which constitute our ‘training’ dataset will be considered in subsequent calculations and analyses of the performance of the interpolated maps. The 24 ‘observed’ profiles will serve as ‘validation’ dataset or ‘ground truth’ for the corresponding validation analyses. However, in building the final BD_{avg} and PSWHC maps for Lower Congo, both 24 and 248 profiles will be included in the mapping process.

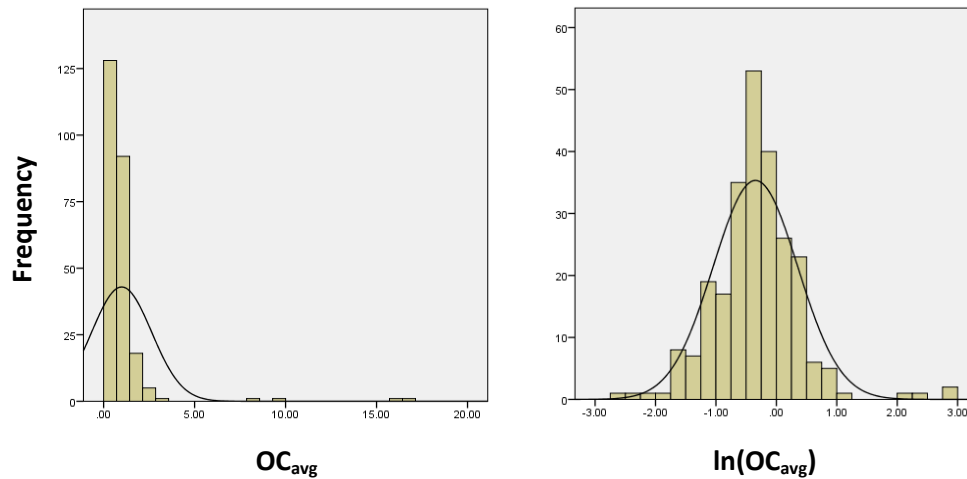


Figure 10-2. Histograms (bars) and theoretical normal distribution (lines) for total OC_{avg} values and log-transformed OC_{avg} values of 248 Lower Congo soil profiles.

Table 10-2. Descriptive statistics of $Clay_{avg}$, $Sand_{avg}$, OC_{avg} , BD_{avg} and PSWHC for Lower Congo soil profiles.[†]

	Mean	Median	Std dev	Min	Max	CV (%)
$Clay_{avg}$ (%)	33.0 (40.2)	32.0 (47.2)	21.0 (26.5)	0.0 (1.6)	84.0 (86.6)	63.6 (66.0)
$Silt_{avg}$ (%)	20.4 ()	18.0 ()	13.2 ()	2.0 ()	65.0 ()	64.9 ()
$Sand_{avg}$ (%)	46.2 (40.8)	45.0 (32.3)	28.7 (31.6)	2.0 (1.6)	97.0 (96.8)	62.2 (77.5)
pH_{avg} (%)	5.2 ()	5.1 ()	0.6 ()	3.3 ()	7.6 ()	12.0 ()
OC_{avg} (%)	0.99 (2.28)	0.70 (1.56)	1.64 (2.10)	0.07 (0.56)	16.52 (8.79)	165.8 (91.9)
BD_{avg} ($Mg\ m^{-3}$)	— (1.38)	— (1.39)	— (0.13)	— (1.08)	— (1.58)	— (9.2)
PSWHC (cm)	— (9.74)	— (9.49)	— (3.30)	— (4.55)	— (21.41)	— (33.9)

[†] The first figures are calculated based on 248 sampling locations of previous soil surveys with no BD_{avg} and PSWHC data and the figures in parentheses are calculated for the 24 new sampling locations with lab-determined BD_{avg} and PSWHC data.

10.3.2 Spatial variability of $Clay_{avg}$, $Silt_{avg}$, $Sand_{avg}$, pH_{avg} and OC_{avg}

Experimental semi-variograms of $Clay_{avg}$, $Sand_{avg}$, $\ln(Silt_{avg})$, $\ln(pH_{avg})$ and $\ln(OC_{avg})$ are depicted in Figure 10-3. $Clay_{avg}$ and $Sand_{avg}$ contents did not exhibit a clear spatial structure and it was not straightforward to fit a theoretical model to the data. This may be partially due to the pronounced variability of these properties as shown by their high CVs. In predicting sand content in Lower Congo based on geostatistical interpolation, Van Meirvenne et al. (1994) showed that stratification into more homogeneous strata could account for an important part of the overall variance. However, these authors warned that, due to clustering of data in some regions of Lower Congo, stratification (zoning) might result into too few observations per stratum to characterize the semi-variogram accurately. This is particularly true for large areas such as the Lower Congo region where the distribution of sampling locations may be considered as erratic for geostatistical applications. However, in soil mapping, the

level of accuracy required depends on the purpose the end-user has in mind. For the log-transformed soil property values, $\ln(\text{pH}_{\text{avg}})$ showed a clear spatial structure compared to $\ln(\text{Silt}_{\text{avg}})$ and $\ln(\text{OC}_{\text{avg}})$.

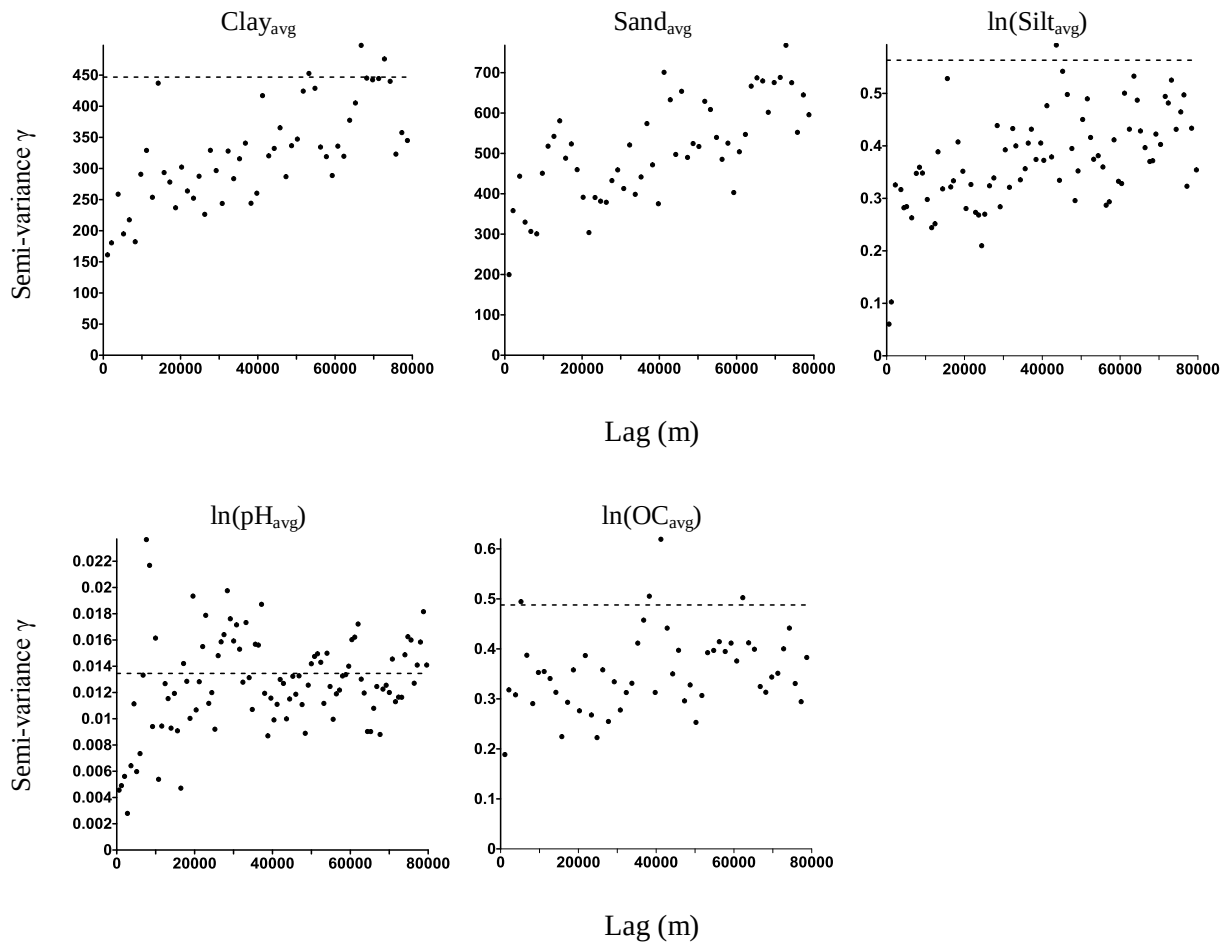


Figure 10–3. Experimental isotropic semi-variograms (black circles) calculated for depth weighted average of clay (Clay_{avg}), sand (Sand_{avg}) and log-transformed depth-weighted average of silt (ln(Silt_{avg})), pH (ln(pH_{avg})) and OC (ln(OC_{avg})) for 248 Lower Congo soil profiles. The dashed lines represent the global semi-variance.

A simplified stratification (zoning) based on the aforementioned regions ('clayey' and 'sandy') was applied on Clay_{avg}, Sand_{avg} and ln(OC_{avg}) data in order to minimize the overall variance across the dataset. The simplified stratification only reduced slightly the overall variance and did not much improve the fit of the experimental semi-variograms. The best improvement was observed for Clay_{avg} interpolated in the clayey region: the nugget effect was reduced from 150 to 100%² and the global variance from 447 to 379%² after stratification as shown in Figure 10–4. For the 'clayey' region, an exponential semi-variogram seems to fit better the data after stratification. The above observations suggest that for soil texture, a good stratification is needed to sensitively reduce the overall variance. In their study, Van Meirvenne et al. (1994) achieved better results for their spatial prediction of clay and sand content in stratifying the Lower Congo region into four strata based on a previous reconnaissance map with soil texture at 60 cm depth as the main stratification criterion. The fact that Clay_{avg} and Sand_{avg} contents are depth-weighted average values could have reduced the spatial dependence of these properties. Since the focus of this study is on spatial variability of hydrophysical soil properties and not on texture, only a stratification of the study area in two 'textural' layers will be considered. Moreover, stratification in more than two strata may result in too small sample size within

some strata to perform acceptable geostatistical analyses. On the other hand, stratification per layer (horizon) might also improve the spatial structure. Santra et al. (2008) stated that most of the studies have dealt with variation of soil properties over two-dimensional surfaces. However, soil properties also vary largely along depth (third dimension) due to active soil-forming processes.

On the other hand, spatial predictions of OC_{avg} or $\ln(OC_{avg})$ would bring along the most uncertainty within the physico-chemical soil properties notably with a high CV of 165.8% for OC_{avg} . A close analysis of the semi-variogram of OC_{avg} revealed the presence of a kind of cyclic pattern generally known as the ‘hole effect’. This spatial pattern of OC was also found by other researchers, e.g., Santra et al. (2008). Stratification and log-transformation of the OC data did not bring any substantial change in the experimental semi-variogram. One of the possible ways of improvement would be the use of a theoretical model which takes into account the ‘hole effect’. This would be particularly useful when continuous maps of carbon stocks are to be produced for the Lower Congo region.

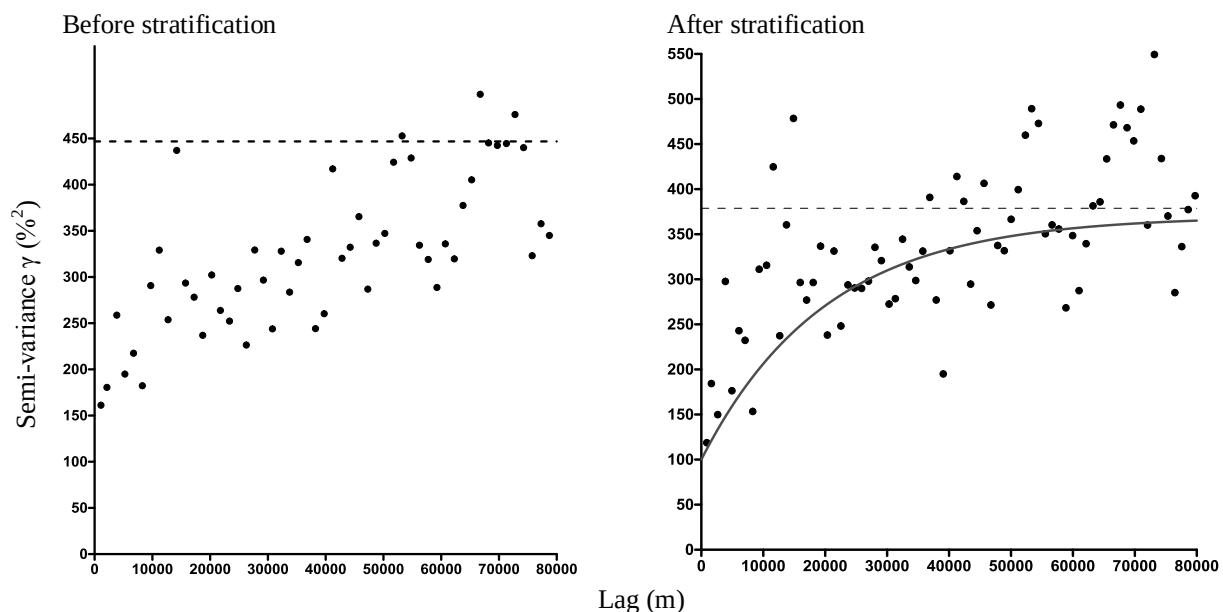


Figure 10–4. Reduction of nugget effect and the global semi-variance (dashed lines) of $Clay_{avg}$ after stratification of the study area in a ‘clayey’ zone and a ‘sandy’ zone. The grey line represents the theoretical semi-variogram (exponential model) fitted to the data located in the ‘clayey’ zone.

However, this goes beyond the scope of this study. Based on the above preliminary results, further investigations should be conducted in order to provide reliable continuous maps of $Clay_{avg}$, $Sand_{avg}$ and OC_{avg} to interested users. As the focus of this study is on mapping BD_{avg} and $PSWHC$ which are PTF-estimated values and not ‘observed’ values as such, only a simplified stratification based on two main strata (‘clayey’ and ‘sandy’) as shown in Figure 10–1 will be considered.

10.3.3 Comparison between ordinary kriging and inverse distance weighting

Application of OK interpolation requires an experimental semi-variogram and a theoretical model to describe the spatial structure of the data. Contrarily to $Clay_{avg}$, $Sand_{avg}$ and OC_{avg} , values of BD_{avg} and $PSWHC$ are derived from PTFs. Therefore, spatial dependencies of estimated hydrophysical properties might differ from one PTF to the other. Wösten et al. (1985) found that the CV of the potential amount of $PSWHC$ is much lower than that of the soil properties used for calculation of the PAW. The same observation was made in the present study (Table 10–2). This indicates that the spatial variability of a given soil property is different when its value at a certain location was directly observed than when it was estimated from, e.g., an equation.

Romano and Santini (1997) wrote that while deviations from actual values of a spatially dependent variable caused by the use of a simplified estimation procedure can have little effect on the calculation of the mean value in a certain area, the spatial distribution of the estimated values can differ from that of the actual values in such a way to alter the patterns of the semi-variograms significantly.

The results for each scenario considered are presented in Table 10–3. From Table 10–3, IDW and OK were found to yield comparable results in terms of RMSD and R^2 , but slightly better results were generally found for IDW. However, one should be aware that the seemingly better performance of the IDW interpolation method was mainly due to clustered sampling locations which are better handled by the IDW algorithm than by OK. When a de-clustering process was performed by excluding from the interpolation process points which are, e.g., within 10,000 m of the target sampling point, the goodness-of-fit (R^2) decreased substantially for the IDW interpolations and the overall error (RMSD) increased. This loss of performance was observed for most of the PTFs selected. On the other hand, OK was robust to the clustering effect, and its R^2 and RMSD values remained practically unchanged. After the de-clustering process, the OK algorithm scored better than IDW for most cases despite the fact that a spherical model was not always found to be the most appropriate model. This shows that OK-generated maps should be of better quality than IDW-generated maps in spatially representing hydrophysical properties of Lower Congo soils.

Table 10-3. Prediction performance of PSWHC values by inverse distance weighting and ordinary kriging for different PTFs' approaches, models and sets of predictors through cross-validation. †

		Inverse distance weighting		Ordinary kriging	
		RMSD	R^2	RMSD	R^2
MLR	Predictors				
	SSC	1.360 (1.590)	0.555 (0.203)	1.390 (1.391)	0.521 (0.521)
	SSCpH	1.435 (1.646)	0.564 (0.284)	1.466 (1.466)	0.529 (0.529)
k-NN	SSC	2.350 (2.330)	0.330 (0.197)	2.297 (2.298)	0.289 (0.289)
	SSCpH	2.052 (2.170)	0.451 (0.248)	2.061 (2.062)	0.411 (0.411)
	SSCpHOC	1.933 (2.082)	0.478 (0.261)	1.961 (1.961)	0.428 (0.428)
PSF	Detailed PSD + BD††	2.557 (2.922)	0.519 (0.140)	2.506 (2.506)	0.527 (0.527)
'Tropical' VB PTFs					
	SSC	3.755 (3.261)	0.156 (0.287)	3.624 (3.792)	0.153 (0.142)
'Temperate' RB PTFs					
	SSC+OM†††	6.805 (6.446)	0.343 (0.327)	6.867 (6.867)	0.324 (0.324)

PTF is pedotransfer function; PSWHC is profile soil water holding capacity (cm/m depth); RMSD is root mean square difference (cm/m depth); R^2 is the coefficient of determination (–); MLR is multiple linear regression; k-NN is k-nearest neighbour; PSF is pore-solid fractal; VB is van den Berg et al. (1997); RB is Rawls and Brakensiek (1982); SSC is texture; OC is organic carbon; OM is organic matter; pH is acidity; PSD is particle size distribution; BD is bulk density.

† Figures between parentheses refer to RMSD and R^2 values found when a de-clustering process was performed by excluding points which are within 10,000 m of the target sampling point in the interpolation procedure;

†† BD=1.37, 1.34 and 1.46 Mg m⁻³ for fine, medium and coarse textured soils, respectively.

††† OM=1.724*OC

Furthermore, several clustered sampling locations enhanced the 'bull's eye' phenomena in most preliminary continuous maps derived from IDW. Generation of 'bull's eyes' surrounding the

observation position within the grid area is recognized in literature as one of the weaknesses of the IDW interpolation method (Isaaks and Srivastava, 1989). On the other hand, OK produced visually appealing maps from irregularly spaced data. This interpolation method attempts to express trends suggested in the data, so that, for example, high points might be connected along a ridge rather than isolated by ‘bull's eye’ type contours. Kriging is a very flexible and robust interpolation method that could better deal with the numerous clustered sampling points than IDW does. Similar observations were also made with BD_{avg} . Clustered points in the Lower Congo map (Figure 10–1) are the results of free soil survey campaigns that took place decades ago. Soil profiles were selected along catenas (from upslope to downslope) or along transects or road cuts to better characterize the soils present in a given area. Some sampling points were located in a relatively short distance and this resulted in a quite irregularly sampling scheme with many clustered points. Based on this brief comparative study, OK was selected as the best interpolation method to produce continuous maps of hydrophysical soil properties in Lower Congo.

10.3.4 Effects of the empirical PTFs-estimated values on the quality of inverse distance weighting and kriging interpolations

According to Table 10–3, differences in IDW and OK predictions between PTFs can be also observed. In the first case, only texture data were considered as input to predict water retention for Lower Congo soils. For IDW and OK interpolations, RMSD values yielded by interpolation of PSHWC estimated by MLR PTFs, are comparable but lower than those found with the k-NN approach. The physico-empirical symmetric PSF model of Hwang et al. (2011) yielded lower prediction performance compared to MLR and k-NN. This was principally due to large PSWHC values estimated by the PSF model compared to values estimated by both the MLR and the k-NN models. The relatively better performance of the MLR equations compared to the kNN and the PSF approaches might be due to the fact that MLR PTFs with only texture as potential predictors could describe better the spatial variability of the estimated PSWHC values. A better fitting of the spherical model to the experimental semi-variogram cloud was achieved with the MLR-estimated values than with the kNN-estimated values.

In the second case, texture, OC and/or pH were considered as predictors. The influence of the PSWHC values estimated by locally developed MLR and k-NN PTFs on the quality of the IDW and OK interpolations were also compared with the ones estimated by two published PTFs: one derived from a ‘tropical’ soil dataset (van den Berg et al., 1997) and one from a ‘temperate’ soil dataset (Rawls and Brakensiek, 1982). The influence of three factors on the interpolation quality of the IDW and OK was investigated here: the effects of the pedotransfer modelling approach, of an additional input (pH and/or OC) and of the geographical origin of the PTF.

When additional predictors such as pH or OC, are added on top of texture to derive PSWHC values, MLR and k-NN exerted a somewhat opposite influence on the interpolation quality of IDW and OK. Addition of new predictors of water retention tended to reduce the interpolation error for the pattern-recognition approach whereas the opposite was observed for the equation-based approach. As shown in Chapter 6, the pattern-recognition approach was able to achieve high prediction accuracy with different sets of predictors. As expected, the published MLR PTFs both yielded higher interpolation errors compared to the locally derived MLR and k-NN PTFs for both IDW and OK. This study shows that different approaches, sets of predictors and pedological origins of PTFs lead to a different assessment of the spatial variation of hydrophysical properties in a given area. Previous researchers (e.g. Romano and Santini, 1997; Nebel et al., 2010) came to the same conclusion using a regular sampling scheme of measured soil water retention points on a very small area (farm scale).

It should be emphasized that the ‘temperate’ PTFs of Rawls and Brakensiek (1982) showed two times higher RMSD values compared to the ‘tropical’ PTFs of van den Berg et al. (1997). This confirms the conclusions of Chapter 5 and of other researchers (e.g. van den Berg et al., 1997; Hodnett and Tomasella, 2002; Tomasella and Hodnett, 2004) that ‘temperate’ PTFs are not suitable for highly weathered soils of the humid tropics. In Chapter 5, large discrepancies were found between measured and predicted water content at -33 kPa and -1500 kPa by the PTFs of Rawls and Brakensiek (1982) when applied on a limited dataset of Lower Congo soils. A clear tendency for overestimation of water retention was observed and this will clearly result in an overestimation of PSWHC. When values generated by these ‘temperate’ PTFs are used in, e.g., mapping agro-ecological zones in Lower Congo, large areas with high PSWHC and therefore good potential for agriculture will be identified. This will lead to an erroneous delineation of agro-ecological zones with a too optimistic appreciation of their suitability in relation with PSWHC. This can seriously bias the judgment of potential land use managers. Therefore, it may be a less risky option to apply a ‘tropical’ PTF than a ‘temperate’ one when deriving agro-ecological land use maps for regions dominated by highly weathered soils and for which no hydraulic information is available. Kar et al. (2004, 2009) showed the importance of hydrophysical soil characteristics and their spatial distribution in defining an action plan for sustainable management of watersheds used for rain-fed rice production in India. Based on the above functional evaluation, the OK algorithm as well as PTFs based on the k-NN approach were selected to produce continuous maps of BD_{avg} and PSWHC for Lower Congo.

10.3.5 Mapping soil bulk density in Lower Congo

A preliminary test showed that a simple averaging of BD irrespective of the thickness of each layer of a soil profile within 100 cm depth (designated here as BD^*_{avg}) exhibited a much better fit of the experimental semi-variogram than a depth-weighted average of BD (BD_{avg}). Therefore, BD^*_{avg} values will be considered in the mapping process. The experimental semi-variograms showed that the BD^*_{avg} values derived from both the MLR and the k-NN models using texture and OC as predictors exhibited a good spatial structure. A spherical model could be easily fitted to the data. However, it was found that better quality could be achieved after stratification of the sampling points into a ‘clayey’ region (204 points) and a sandy region (44 points). However, due to the limited number of sampling points in the ‘sandy’ region (only 44 points) and the irregularity of the sampling scheme, particularly in the oriental region (Figure 10–1), the associated experimental semi-variogram showed an erratic behaviour that could not be modelled. Therefore, no stratification was applied in the production of the final map of BD^*_{avg} . An independent test was performed to derive the most suitable map based on BD^*_{avg} values of the 24 profiles with lab-determined BD. It was found that the map of k-NN derived values yielded a slightly better prediction quality ($MD = -0.03 \text{ Mg m}^{-3}$, $RMSD = 0.11 \text{ Mg m}^{-3}$ and $R^2 = 0.453$) than the one of MLR derived values ($MD = -0.05 \text{ Mg m}^{-3}$, $RMSD = 0.12 \text{ Mg m}^{-3}$ and $R^2 = 0.384$). Therefore, a final map including the 24 observed points was constructed based on the k-NN approach as depicted in Figure 10–5 with its associated spherical semi-variogram (Figure 10–6).

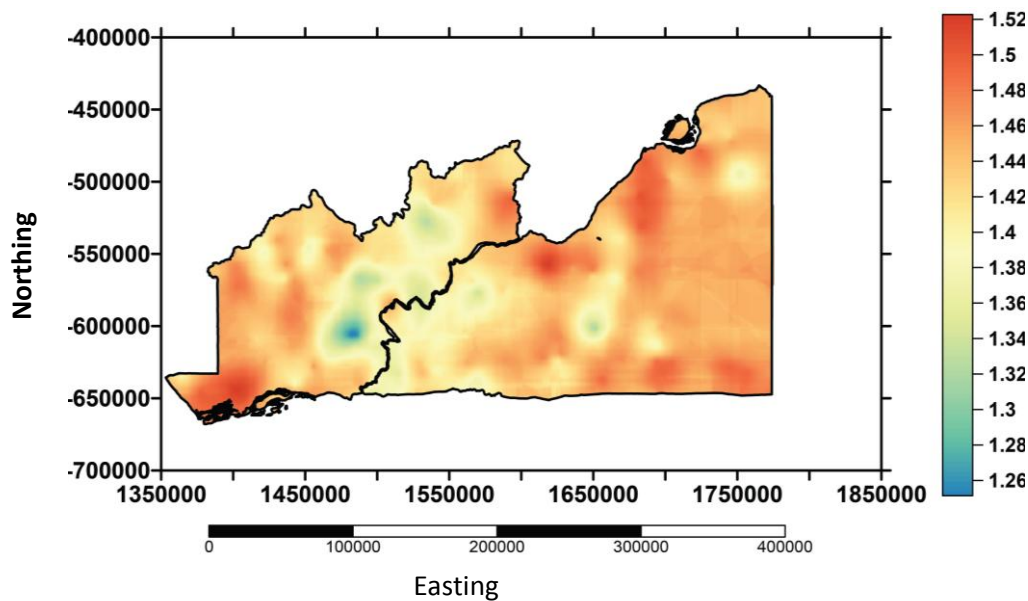


Figure 10–5. Spatial variability of bulk density (g cm^{-3} or Mg m^{-3}) of soils in Lower Congo

Despite the non-stratification of the study area, the coastal and the oriental regions dominated by coarse textured soils showed the highest BD^*_{avg} values compared to some areas in the axial region dominated by high clay content. This BD map combined with OC content information can serve to derive, e.g., a carbon stock map for Lower Congo at a finer resolution (1:500,000) than the existing map derived by Batjes (2008) at 1:1,000,000 scale and based on SOTER data and pedotransfer rules. This refinement on the spatial distribution of OC stocks in this region will be useful to studies related to climate change in Central Africa.

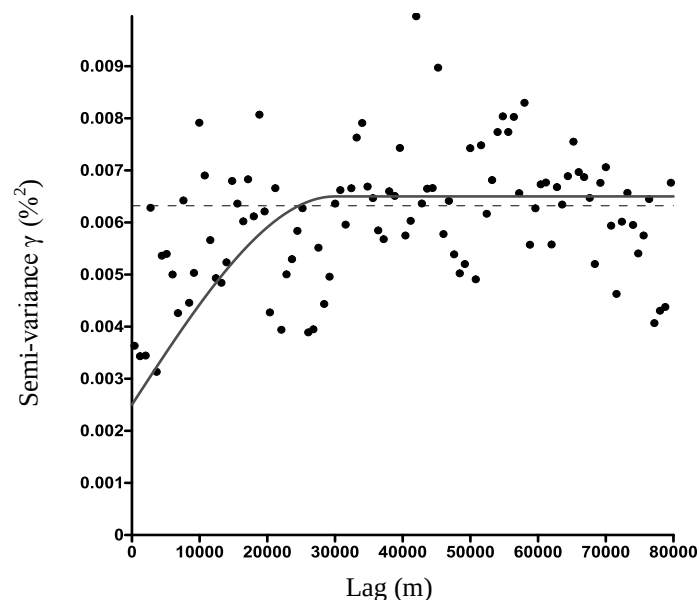


Figure 10–6. Experimental and theoretical isotropic semi-variograms (spherical model) for bulk density of Lower Congo soils estimated by the k-nearest neighbour approach with clay, silt, sand and OC as predictors. Black circles represent the experimental semi-variance and the grey line represents the theoretical semi-variance).

10.3.6 Mapping PSWHC in Lower Congo

Similar to BD, an independent test was performed to assess the predictive quality of PSWHC maps derived from estimated values yielded by equation-based (MLR), pattern recognition based (k-NN) and fractal based (PSF) models. Maps were constructed with the OK interpolation method using the best theoretical semi-variogram model. Spherical, exponential and linear models were fitted to the corresponding data. The predictive quality of the derived maps was tested using the 24 sampling locations with ‘observed’ PSWHC data. A simplified stratification similar to the BD map was performed, but this did not improve the predictive quality of the resulting maps. This may be partially due to the inherent properties of the highly weathered soils related to water retention. Despite their high clay content, these soils have physical properties similar to sandy soils (van den Berg et al., 1997; Hodnett and Tomasella, 2002; Tomasella and Hodnett, 2004). Therefore, stratification into ‘clayey’ and ‘sandy’ zones may not be discriminating enough to spatially characterize the water retention properties of Lower Congo soils. Prediction performance of each map in terms of MD, RMSD and R^2 is presented in Table 10–4.

The results showed that the best prediction quality was achieved with PSWHC values derived from the k-NN approach with texture and pH as predictors. On the other hand, it appears that using OC as predictor in the k-NN approach contributed somewhat to a decrease in the overall prediction error. This may be partially attributed to the irregular spatial behaviour of this soil property.

Table 10-4. Prediction performance of each derived map in terms of MD, RMSD and R^2 for Lower Congo PTF-estimated PSWHC values.

		Theoretical model	MD (cm/m depth)	RMSD (cm/m depth)	R^2 (–)
MLR	SSC	Spherical	1.10	3.06	0.276
	SSC + pH	Spherical	1.13	3.00	0.557
k-NN	SSC	Linear	1.35	3.38	0.131
	SSC + pH	Exponential	0.58	2.91	0.583
	SSC + pH + OC	Linear	0.12	2.97	0.539
PSF					
	Detailed PSD+ BD†	Linear	4.41	5.43	0.087

PTF is pedotransfer function; PSWHC is profile soil water holding capacity; MD is mean difference (cm/m depth); RMSD is root mean square difference (cm/m depth); R^2 is the coefficient of determination (–); MLR is multiple linear regression; k-NN is k-nearest neighbour; PSF is pore-solid fractal; SSC is texture; OC is organic carbon; pH is acidity; PSD is particle size distribution; BD is bulk density.

† BD = 1.37, 1.34 and 1.46 Mg m^{–3} for fine, medium and coarse textured soils, respectively.

A map based on PSWHC values estimated by the k-NN approach with clay, silt, sand and pH as predictors was eventually selected to reflect spatial variability of the water retention characteristics using 248 estimated PSWHC values and 24 ‘observed’ values (Figure 10–7). The corresponding exponential semi-variogram is depicted in Figure 10–8.

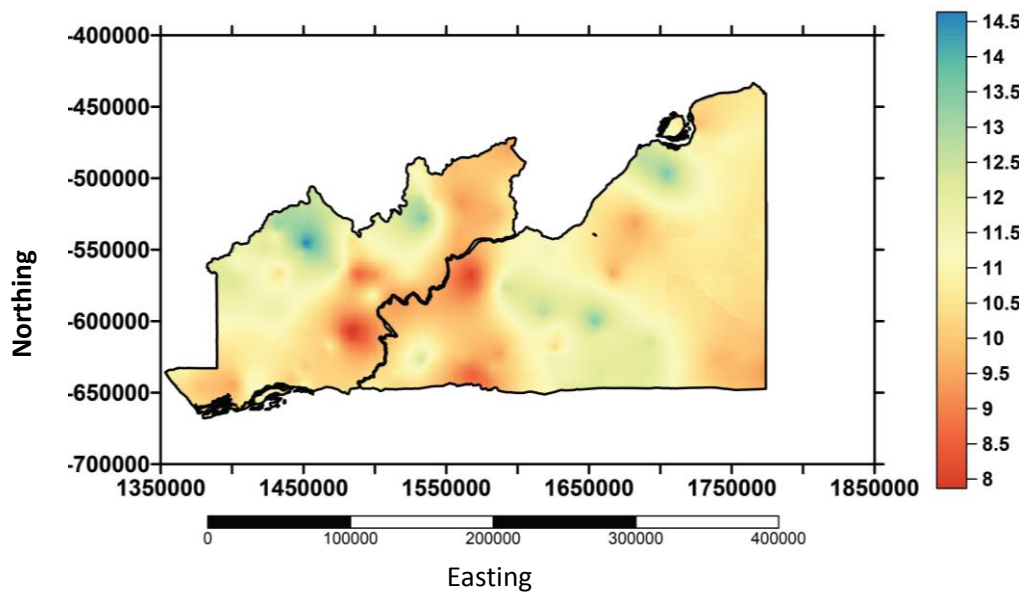


Figure 10–7. Spatial variability of profile water holding capacity (cm per m depth) of soils in Lower Congo

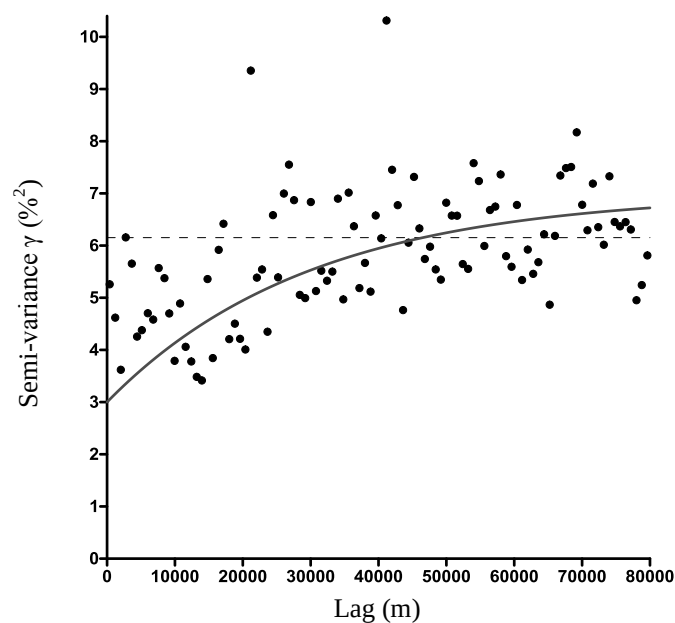


Figure 10–8. Experimental and theoretical isotropic semi-variograms (exponential model) for profile soil water holding capacity of Lower Congo soils estimated by the k-nearest neighbour approach with clay, silt, sand and pH as predictors. Black circles represent the experimental semi-variance and the grey line represents the theoretical semi-variance.

This map shows high spatial variability in PSWHC of soils in Lower Congo. Large zones showing low PSWHC (< 12 cm per m depth) both in the ‘sandy’ as well as in the ‘clayey’ regions of Lower Congo can be identified. This illustrates that the Lower Congo region is dominated by highly weathered and coarse textured soils with low PSWHC. This is in agreement with the findings of other researchers about the specific hydraulic behaviour of highly weathered soils of the humid tropics compared to soils from temperate regions (van den Berg et al., 1997; Tomasella et al., 2000; Medina et al., 2002; Hodnett and Tomasella, 2002). Ferralsols and other related soils (Acrisols and Nitisols) are known to have low water content both at FC and at PWP resulting in a low PSWHC. Some restricted areas in

Lower Congo were found to have higher PSWHC (> 13 cm per m depth). Soils of these regions are soils with high OC content due to agricultural activities and/or topographical location (lowland soils) or with high silt content. This is particularly the case with soils of Menkao, a sandy area in the northern part of the oriental region (Kinshasa), which are known to have high OC content.

This hydraulic Lower Congo map indicates that a better management of soil and water resources related to agricultural activities should take into account the relatively low PSWHC of soils located in this region. Plant available water capacity is one of the main determining factors for growing crops. Therefore, reliable spatial information concerning soil hydraulic characteristics are important for simulating the soil water balance and estimating crop yields under various management scenarios subjected to different rainfall regimes. The crop cycle of common crops cultivated can be fitted into the growing period taking into account climate-, soil- and crop-specific sowing dates avoiding water stress conditions as much as possible.

10.4 Conclusion

A simple functional evaluation of BD and water retention PTFs derived for Lower Congo soils along with some published PTFs has been performed. This evaluation study concerned the application of PTFs in describing the spatial variability of hydrophysical properties of highly weathered soils and coarse textured soils which are dominant in this region. To produce continuous maps of these hydrophysical properties, OK was found to be a better interpolation method compared to IDW. On the other hand, 'temperate' PTFs tested showed a strong tendency to overestimate the PSWHC of Lower Congo soils and should not be used to derive continuous maps of hydrophysical properties of highly weathered soils. After testing some pedotransfer modelling approaches and predictive models with different sets of predictors, the k-NN method with texture and OC as predictors of BD^*_{avg} and the k-NN method with texture and pH as predictors for PSWHC were selected as the best options for mapping hydrophysical soil properties in the study area. Subsequently, a BD map and a SHWC map for the Lower Congo region were derived. Both maps at the scale of 1:500,000 showed considerable spatial variability of soil hydrophysical characteristics that should be taken into account in agricultural and environmental studies in this region. In a more advanced step, BD and SHWC maps can be combined with various thematic maps related to climate, geology, mineralogy, topography, hydrology, vegetation, agriculture and other useful data into a geographical information system. This will result in an integrated land use map that will be a tool of choice in the hands of land managers and other end-users to promote sustainable management of natural resources of Lower Congo.

11 General discussion, conclusions and future research

11.1 General discussion

McBratney et al. (2002) reported that the estimation of soil water retention constitutes the most comprehensive research topic in formulation of pedotransfer functions (PTFs). Since the publication of the last review paper of Tomasella and Hodnett (2004), exclusively dedicated to published PTFs to predict water retention of tropical soils was published, the need was growing to provide an up-to-date repository of past and recent published articles as well as proceeding papers dealing with water retention PTFs for soils of the humid tropics. Another important objective of this review was to categorize published water retention PTFs based on various approaches and list their application to soils of the humid tropics based on available literature data. This allowed identifying the most common and less common approaches used to estimate water retention of soils in tropical regions. From this review study conducted in Chapter 3, it was found that among the PTFs developed for soils in the humid tropics:

- 91% of the PTFs are based on an empirical approach and 9% on a semi-physical one;
- 97% of the empirical PTFs were derived based on multiple linear regression (MLR) and polynomial of n^{th} order techniques and 3% on the k-nearest neighbour (k-NN) approach;
- 84% of the continuous PTFs are point-based, 16% are parameter-based and 0% are pseudo-continuous PTFs;
- 97% of the continuous PTFs are equation-based and 3% are based on pattern-recognition; and
- 26% of the ‘tropical’ water retention PTFs were developed for soils in Brazil, 26% for soils in India, 11% for soils in other countries in America (US, Mexico and Cuba), 11% for soils in Africa (Nigeria, Sierra Leone, Tanzania, Uganda and Zambia).

Most of the local ‘tropical’ PTFs published in peer-reviewed journals originated from Brazil, Nigeria and India. In other regions of the humid tropics, e.g., Central Africa, there were no (recorded) efforts in the development of local water retention PTFs.

The above figures confirm the earlier statement of Schaap (2005) that ‘with the exception of a few studies, hydraulic data and corresponding indirect methods about tropical soils are a virtual *terra incognita*’. This situation has not changed much until present.

Regarding predictor variables used, particle size distribution (PSD) or texture, organic carbon/organic matter (OC/OM) and bulk density (BD) were found to be the most common in tropical PTFs, whereas other soil properties that are found in various databases of soils from the humid tropics such as pH, cation exchange capacity (CEC), dithionite-citrate-bicarbonate extractable iron (DCB-Fe) and aluminium (DCB-Al), were considered to a limited extent only.

In the review study conducted in Chapter 3 on PTFs to predict water retention of soils in the humid tropics, PSD was found to be the most common predictor in ‘tropical’ PTFs. As indicated by Batjes (1996) and Nemes et al. (1999), lack of reliable methods for particle-size standardisation prevented international soil databases to include data that did not comply with the FAO/USDA particle size description. Therefore, a good standardisation of PSD is a prerequisite to achieve compatibility of soil data among various tropical countries with different texture classification systems. A direct benefit of particle size standardisation resides in the current and future development of multi-scale PTFs as estimators of soil hydrophysical properties.

In literature, several mathematical models have been proposed to represent PSD. Numerous studies evaluated the performance of such models to describe PSD of soils from temperate areas and only few

research

attempts were made to assess the fitting and prediction performance of PSD models when applied on soils of tropical areas.

In Chapter 4, the performances of two-, three- and four-parameter models for describing cumulative PSD of soils of the humid tropics based on a large dataset from Lower Congo were evaluated. We found that several models like the two and three-parameter models of Haverkamp and Parlange (1986) (VG_2p and VG_3p), the three-parameter model of Skaggs et al. (2001) (Sk_3p) and the four-parameter equation of Gompertz tested in Nemes et al. (1999) (G_4p) were not suitable to describe PSD of soils in the humid tropics. On the other hand, the two and three-parameter models of Fredlund et al. (2000) (F_3p and F_4p), the three-parameter model of Weibull used by Assouline et al. (1998) (W_3p) and the four-parameter function of Andersson's reported by Jauhiainen (2004) (A_4p) showed outstanding fitting performance. The fitting performance of PSD models was also dependent of the textural class, the broad textural group and the bimodal character of the soil. For the most frequent, textural classes in the dataset (i.e. C, CL, S, SCL and SL), the F_3p and A_4p models were the best closely followed by the W_3p model. While the F_3p model performed better than the A_4p model for coarse textured soils, the opposite was observed for fine textured soils. The W_3p model showed acceptable fitting performance for fine, medium and coarse textured soils. The performance of the PSD models was found to be better for bimodal soils, which are common in the humid tropics, than for unimodal soils. We conclude that F_3p, F_4p, W_3p and A_4p models are highly recommended in order to get a better description of the PSD of soils of the humid tropics and achieve standardisation of PSD in international tropical databases. These PSD models can be particularly useful when assessing the predictive ability of various PTFs developed based on soil datasets with different texture classification systems (e.g. different definition of clay, silt or sand fraction) than the test datasets.

Before engaging in developing specific water retention PTFs for Lower Congo soils, the ability of some 'point-based' and 'parameter-based' PTFs from various geographical origins to predict water content at -33 kPa and -1500 kPa was statistically assessed based on a limited soil dataset (60 samples) from Lower Congo (Chapter 5). This further allowed to identify potential predictors of water retention of highly weathered soils. A number of published and well-known PTFs derived from soils of temperate-climate regions ('temperate' PTFs) and from soils of the sub(humid) tropics ('tropical' PTFs) were selected. Based on various statistical measures, it was proved that there is no one single set of PTFs that is valid for soils of Lower Congo. The point-based 'tropical' PTFs of Oliveira et al. (2002) and Dijkerman (1988) performed best at -33 kPa, while those of Arruda et al. (1987) and Pidgeon (1972), also developed on a 'tropical' soil database were best at -1500 kPa. Regarding the parameter-based PTFs which predicted the van Genuchten (1980) parameters, the 'tropical' PTFs of Hodnett and Tomasella (2002) and the 'temperate' PTFs of Schaap et al. (2001) gave the best results. On the other hand, widely applied 'temperate' PTFs of Gupta and Larson (1979) and Rawls and Brakensiek (1982) showed poor predictive performance with a great tendency to overestimate water retention of Lower Congo soils.

These results suggested that estimates of water retention by some 'temperate' as well as 'tropical' PTFs might induce errors in the outputs of watershed models used in various agricultural studies under the humid tropics. Large discrepancies in the derived soil hydraulic data can substantially reduce the quality of the modelling results particularly in regions where soils may have been formed and evolved in similar climatological and pedological conditions as soils from Lower Congo.

This prospective study revealed that there is a need to develop new water retention PTFs based on a larger dataset of soils from Lower Congo. Moreover, PTFs for predicting BD and K of these soils will

be very useful as such crucial information is generally missing in the D.R. Congo soil datasets. From Chapter 3 and Chapter 5, nearly all the investigated ‘tropical’ PTFs were found to be equation-based. However, other approaches which are generally termed as nonparametric i.e. not based on equations are gaining popularity in various soil-related fields.

Nonparametric approaches such as the k-NN method are considered attractive for pedotransfer modelling in hydrology. However, nonparametric approaches have not been applied so far to predict water retention of highly weathered soils in the humid tropics. A variant of the k-NN algorithm developed by Nemes et al. (2006a) was developed and tested to predict the water retention of soils from Lower Congo based on an international dataset (IGBP-Trop) of soils of the (sub)humid tropics. Two design parameters, k and p , that are user-defined and determined before and independent of applying the nonparametric k-NN algorithm were optimized to take better advantage of the k-NN variant introduced in this study. The optimized k and p values were found to be similar to those of previous studies on ‘temperate’ soils.

The results showed that this k-NN variant was able to estimate water content at eight different matric potentials (0, -1, -3, -10, -20, -50, -250, and -1500 kPa), i.e., from the wet to the dry range of the soil water retention curve (SWRC), with an average root mean square difference, $\text{RMSD} < 0.046 \text{ m}^3 \text{ m}^{-3}$ when $\text{SSC} + \text{BD}$ or $\text{SSC} + \text{BD} + \text{CEC}$ were selected as input variables. Particularly with the best combination of input variables ($\text{SSC} + \text{BD} + \text{CEC}$), the overall error in predicting water retention of 139 selected Lower Congo soils was remarkably low: 0.0360 to 0.0390 $\text{m}^3 \text{ m}^{-3}$ in the dry and very wet ranges and 0.0490 to 0.0510 $\text{m}^3 \text{ m}^{-3}$ in the intermediate range (i.e., -3 to -50 kPa) of the SWRC, as compared to values reported in literature for other PTFs.

The overall prediction performance of the proposed nonparametric approach was compared with two tropical equation-based PTFs of Hodnett and Tomasella (2002) and Minasny and Hartemink (2011) based on the MLR approach and also using a tropical subset of the IGBP-DIS database with different number of soil samples. The results suggest that the k-NN approach shows comparable prediction performance to the examined MLR PTFs, which makes it a competitive alternative to those equation-based PTFs that are currently available to predict the water retention of soils in the humid tropics. Furthermore, the k-NN variant developed in this study presents, as added benefits, the possibility to estimate prediction uncertainty and the flexibility to incorporate new data without the need to redevelop new equations. This is highly beneficial for developing countries under the tropics where soil databases for agricultural planning are scarce at present, though they are slowly being developed. Additionally, similar prediction performance could be attained with our k-NN variant using less predictors (texture and BD) than the parameter-based PTF of Hodnett and Tomasella (2002) which requires texture, BD, OC, and CEC to be applied. Therefore, PTFs based on the classical MLR approach and on the recently tested k-NN approach (Chapter 6) can be developed to predict hydrophysical properties of Lower Congo soils.

Bulk density (BD) is a key soil property for sustainable land management and carbon stocks assessment. However, data on BD are often missing in various soil survey reports and most soil databases of Central Africa. A set of four ‘tropical’ and four ‘temperate’ published PTFs yielded poor prediction performance with a great tendency to underestimate BD of Lower Congo soils when tested on 196 samples. It was shown that the poor prediction performance of existing BD PTFs cannot only be attributed to the geographical location of the soils in the development dataset. Most training datasets used to derive the published PTFs were dominated by soils rich in OM which contributed to

research

a lower BD. Another reason of the clear underestimation of BD by the tested PTFs could be that some of these PTFs were developed based exclusively on relatively low BD forest soils with a limited number of samples from subsoils whereas in this study, a large proportion of non-forest soils and subsoil samples with relatively higher BD were included. Therefore, these published PTFs were found to be impractical for the Lower Congo region and new PTFs needed to be developed.

To do so, two approaches, namely an equation-based (MLR) and a pattern-recognition approach (k-NN) were tested with different levels of information on the following soil characteristics: texture, OC, pH, CEC, DCB-Fe and DCB-Al. With the same set of predictors, the k-NN method had a better prediction performance than the MLR method with an overall prediction error (in terms of root mean square difference, RMSD) for k-NN of 0.096 Mg m^{-3} when using clay, sand and OC content, and 0.108 Mg m^{-3} with OC content, and DCB-Fe and DCB-Al, compared to the MLR approach with RMSD of 0.137 and 0.126 Mg m^{-3} , respectively. As explained earlier, the k-NN method presents other additional advantages such as flexibility and uncertainty estimation. Furthermore, the results in terms of RMSD values of both approaches were better than those of other published PTFs developed based on temperate and tropical soil datasets. On the other hand, this study demonstrated the relative importance of DCB-Fe and DCB-Al along with other commonly used soil characteristics such as OC and texture in the improvement of BD prediction in both MLR and k-NN approaches. However, it should be kept in mind that BD is a physical soil property that is relatively difficult to predict accurately from commonly used soil characteristics, even with local PTFs due notably to its high spatial variability.

Similarly as for developing hierarchical BD PTFs using a Lower Congo dataset (Chapter 7), different levels of availability of information on soil characteristics were considered to develop local soil water retention PTFs, i.e. based on the Lower Congo hydrophysical dataset established within this study (Chapter 8). The variables considered were texture, BD, OC, pH, CEC, DCB-Fe and DCB-Al and again two approaches were tested. In the first approach, PTFs in the form of MLR equations were formulated to predict water content at eight different matric potentials. In the second approach, a variant of the k-NN algorithm developed previously was used. A comparative analysis of the MLR and the k-NN approaches showed that the k-NN method had a high prediction performance with an overall prediction error (RMSD) $< 0.035 \text{ m}^3 \text{ m}^{-3}$ compared to the MLR PTFs with RMSD $< 0.046 \text{ m}^3 \text{ m}^{-3}$. Furthermore, as expected, the locally developed Lower Congo PTFs performed better than two other published 'tropical' PTFs namely the parameter-based PTFs of Hodnett and Tomasella (2002) and the point-based PTFs of Minasny and Hartemink (2011). On the other hand, the relative importance of soil properties, like DCB-Fe and DCB-Al, in the improvement of prediction of water retention in both approaches was demonstrated along with other more commonly used properties such as BD and CEC.

This study also showed that the k-NN method can be considered as a competitive alternative to more classical MLR PTFs due to the accuracy of the water retention estimation and its flexibility to incorporate new data without the burden to redevelop new equations. The derived PTFs can be used to provide estimates of water retention at FC and the PWP, or to construct a complete SWRC. These hydraulic parameters are used as inputs in various simulation models or, e.g., to derive soil hydraulic property maps.

Pedotransfer functions tested and developed in the preceding chapters (Chapters 5, 6, 7 and 8) are based on the empirical approach. As seen in Chapter 3, few water retention PTFs were based on the

semi-physical approach. In characterization of soil water retention by PTFs, a physico-empirical model based on the pore-solid fractal (PSF) theory has recently gained increasing attention due notably to the growing interest for the fractal approach in soil science. Previous studies related to PSF models for water retention focused on soils from temperate climate regions. Availability of detailed PSD and SWRC data from tropical soil datasets as well as the particular hydraulic behaviour of soils of the humid tropics opened opportunities to investigate the applicability of the PSF approach to describe and predict water retention of highly weathered soils. It was found that the symmetric PSF model developed by Hwang et al. (2011) can relatively well describe the water retention function ($0.010 \text{ m}^3 \text{ m}^{-3} < \text{RMSD} < 0.064 \text{ m}^3 \text{ m}^{-3}$) and performs by far better compared to the asymmetric form of the model ($0.039 \text{ m}^3 \text{ m}^{-3} < \text{RMSD} < 0.092 \text{ m}^3 \text{ m}^{-3}$). This actually contrasted with the observations for Korean soils, that the asymmetric model outperformed the symmetric model except for coarse textured soils. However, the symmetric PSF model seemed to have, compared with our PTFs developed in Chapters 6 and 8, relatively limited prediction ability ($0.035 \text{ m}^3 \text{ m}^{-3} < \text{RMSD} < 0.087 \text{ m}^3 \text{ m}^{-3}$) particularly when the range of soil properties such as OM in the calibration dataset ($1.04\% \pm 1.274$) is relatively different from the validation dataset ($2.26\% \pm 2.507$). Moreover, water retention of fine textured soils with high clay content was found to be more challenging to predict than for medium and coarse textured soils due, among others, to the complexity of the pore geometry in clayey soils. More studies with other soil datasets are required in order to better appreciate the potential of the PSF theory in describing and predicting SWRC of highly weathered soils.

Legacy data and data recently collected data on soils of Lower Congo were used to evaluate PTFs' outputs for practical application. An evaluation of PTFs' outputs for practical application (functional evaluation) has been conducted. This study assessed the ability of the PTFs tested and developed in the preceding chapters to describe the spatial variability of hydrophysical properties of highly weathered soils and some coarse textured soils of Lower Congo. First, the effects of PTF-estimated values on the prediction performance of two interpolation methods (inverse distance weighting, IDW and ordinary kriging, OK) were investigated. The results showed that OK was a better interpolation method compared to IDW. On the other hand, 'temperate' PTFs tested showed a strong tendency to overestimate soil water holding capacity (SWHC) and should not be used for mapping hydraulic properties of highly weathered soils. Secondly, some pedotransfer modelling approaches and predictive models with different sets of predictors were evaluated. This resulted in the selection of the k-NN method with texture and OC as predictors of BD and the k-NN method with texture and pH as predictors of SWHC as the best options for mapping hydrophysical soil properties of Lower Congo soils. Based on these findings, a BD map and a SWHC map for the Lower Congo region were derived at the scale of 1:500,000. Considerable spatial variability of soil hydrophysical characteristics could be observed in both maps. This spatial variability should be taken into account in various agricultural and environmental studies in order to promote rational management of the natural resources of the Lower Congo region.

11.2 Conclusion

The main goal of this study was to develop indirect methods such as PTFs to predict hydrophysical properties of soils in Lower Congo, the South-western part of the D.R. Congo. A thorough review of existing PTFs developed to predict water retention of soils in the humid tropics shows that various novel data-mining techniques such as Artificial Neural Networks, Regression Trees, Support Vector Machines, Genetic Programming as well as pseudo-continuous and ensemble PTFs have not yet been applied for water retention prediction of soils from the humid tropics. Minasny and Hartemink (2011)

research

indicated that with the emergence of available large soil databases in tropical regions, the development of new PTFs based on data mining tools should be boosted.

Various PSD models were tested to provide reliable methods to achieve compatibility between tropical soil databases with different textural classification systems. A comparable study can also be conducted for salt-affected soils of the dry (sub)tropics or alluvial soils which have different properties than highly weathered soils of the humid tropics. Another interesting research topic could be to evaluate the utility of various PSD models to describe cumulative PSD of ‘tropical’ soils that has been determined by laser-diffraction methods.

A prospective study was conducted to evaluate the ability of a number of published point-based and parameter-based PTFs to predict water retention of soils in Lower Congo. Preliminary results of this evaluation study suggest that estimates of water content by several existing ‘temperate’ as well as ‘tropical’ PTFs may induce errors in the outputs of watershed models used in various agricultural studies under the humid tropics. A more advanced step could be the quantification of the uncertainty related to a blind use of water retention estimates provided by such PTFs on the outcomes of soil water management models (e.g., SPAW, AquaCrop) when applied in the humid tropics.

Classical equation-based approaches as well as novel statistical approaches based on pattern-recognition were used to develop PTFs to predict BD, SWRC and hydraulic conductivity (K) of highly weathered soils in Lower Congo. The application of promising pattern-recognition methods such as k-NN should be applied on soils located in the tropics but with different physical, chemical and mineralogical properties. It will be particularly interesting to test the ability of the k-NN technique to predict water retention of other soils found in the humid tropics such as volcanic soils (Andosols), swelling-shrinking soils (Vertisols) or alluvial soils (Fluvisols). The latter present some specific properties and may need completely different reference or training datasets than the ones used in this study in order to provide reliable estimates of their hydrophysical characteristics.

The results of this research showed the potential of other soil characteristics, such as DCB-Fe and DCB-Al to improve prediction of BD and water retention of highly weathered soils. However, further investigations with extended and diversified soil datasets as well as with more sophisticated analytical tools such as mercury intrusion porosimetry, environmental scanning electron microscopy or computer tomography should be conducted to better appreciate the active role of Fe and Al oxides in relation with BD and water retention characteristics of strongly weathered soils. Moreover, other soil information, like categorical soil structural data based on profile description should be investigated as potential predictors of BD and water retention of soils in the humid tropics.

In this study, various hierarchical PTFs were developed to predict BD and water retention for soils of Lower Congo with relatively high accuracy. Notwithstanding its high pedological diversity, Lower Congo does not encompass all the soil types found in the D.R. Congo. Therefore, other PTFs based on various approaches should also be tested and/or formulated particularly in areas where soils may have been formed and evolved in different climatological and pedological conditions than soils from Lower Congo.

To date, studies related to prediction of the SWRC with PSF models of soil structure were limited to soils of temperate regions. Results from the first evaluation study on the fitting and predictive ability of PSF models when applied on Lower Congo soils were not conclusive and showed the need to conduct more investigations in order to better evaluate and/or improve the potential of the PSF theory to explain water retention in highly weathered soils.

The potential user can choose which pedotransfer modelling approach to apply to estimate soil BD and water retention. The MLR approach seems to be the most convenient to apply in order to get a quick

estimation. However, pattern-recognition techniques such as the k-NN method offer various benefits in terms of flexibility and uncertainty estimations. When implemented in a user-friendly software, they can become easy to apply. The application will be provided to the public upon request in Matlab® and Python® programming language at the Soil Physics Research Unit of Ghent University (Belgium).

11.3 Future research

Research on physical and hydraulic aspects of soils in tropical regions is still quite limited. Discussing the role of soil science in agricultural development in East Africa, Muchena and Kiome (1995) concluded that despite the activities of several foreign experts, there is still inadequate expertise in some key disciplines such as soil physics, land evaluation and water management. This situation has not improved that much today. The new research field of hydropedology carries the potential of bridging hydrology and pedology to respond to the lack of data on laborious and expensively determined hydraulic soil properties in temperate as well as in tropical regions. The emergence of initiatives such as the Globally Integrated-Africa Soil Information Service (AfSIS) project resulting in the formation of a newly established comprehensive international soil database for Sub-Saharan Africa will certainly provide new opportunities for interdisciplinary soil and water-related research for, e.g., hydrological catchment modelling or assessment of agricultural production (Wösten et al., 2013).

As mentioned earlier hydraulic conductivity (K) is one of the important hydraulic properties used in various models to simulate variably saturated water flow and solute transport. As indicated by various researchers (Vereecken, 1995; Revil and Cathles, 1999; Minasny and McBratney, 2000; Doussan and Ruy, 2009; Wösten et al., 2013), saturated and unsaturated soil hydraulic conductivity ($K(h)$ and K_{sat}) are arduous, expensive and time-consuming to measure in the field. They are known to vary over several orders of magnitude and have high spatial variability. Therefore, a significant number of field measurements and samples are needed to develop sound-based PTFs for K_{sat} and $K(h)$. Compared to water retention, few PTFs have been dedicated to prediction of these soil properties in literature. This is particularly true for soils in Sub-Saharan Africa for which few data on K_{sat} exist and are often inconsistent with texture of the soil horizon (Wösten et al., 2013). Seeking for more robust and efficient methods to characterize K for real-world applications, various researchers such as Kool and Parker (1988), van Dam et al. (1994), Simunek et al. (1999) used another indirect method called inverse modelling. In this research, PTFs to predict unsaturated and saturated hydraulic conductivity ($K(h)$ and K_{sat}) have not been developed for Lower Congo soils due to time and financial constraints to obtain a statistically significant number of data for this purpose. However, a limited number of field measurements of $K(h)$ with tension disc infiltrometer were performed in autumn 2009 during the sampling campaign. Both indirect approaches (pedotransfer modelling and inverse modelling) need to be investigated to better estimate $K(h)$ and K_{sat} for ‘tropical’ soils.

As part of the AfSIS initiative, a database of soil hydrophysical properties for Sub-Saharan Africa similar to the EU-HYDI database for Europe should be created. Such comprehensive hydropedological database will carry outstanding potential for enhancing the knowledge on ‘tropical’ soils for developing and applying models related to soil water on a pan-African scale.

References

- Adams, W.A., 1973. The effect of organic matter on the bulk and true densities of some uncultivated Podzolic soils. *European Journal of Soil Science* 24, 10-17.
- Adhikary, P.P., Chakraborty, D., Kalra, N., Sachdev, C.B., Patra, A.K., Kumar, S., Tomar, R.K., Chandna, P., Raghav, D., Agrawal, K., Sehgal, M., 2008. Pedotransfer functions for predicting the hydraulic properties of Indian soils. *Australian Journal of Soil Research* 46, 476-484.
- Agyare, W.A., Park, S.J., Vlek, P.L.G., 2007. Artificial neural network estimation of saturated hydraulic conductivity. *Vadose Zone Journal* 6, 423-431.
- Aina, P.O., Periaswamy, S.P., 1985. Estimating available water-holding capacity of western nigerian soils from soil texture and bulk-density, using core and sieved samples. *Soil Science* 140, 55-58.
- Akaike, H., 1973. Information theory and an extension of maximum likelihood principle. In: Petrov, B.N., Csáki, F. (Eds.), *Second International Symposium on Information Theory*. Akadémia Kiado', Budapest, pp. 267-281.
- Al Majou, H., Bruand, A., Duval, O., 2008. The use of in situ volumetric water content at field capacity to improve the prediction of soil water retention properties. *Canadian Journal of Soil Science* 88, 533-541.
- Alexander, E.B., 1980. Bulk densities of California soils in relation to other soil properties. *Soil Science Society of America Journal* 44, 689-692.
- Allen, B.L., Hajek, B.F., 1989. Mineral occurrence in soil environments. In: Dixon, J.B, Weed, S.B. (Eds.), *Minerals in soil environments*. Second Edition. SSSA, Madison, Wisconsin, pp. 199-278.
- Anand, R.R., Gilkes, R.J., 1987. Iron-oxides in lateritic soils from Western-Australia. *Journal of Soil Science* 38, 607-622.
- Andersson, S., 1990. Markfysikaliska undersökningar i odlad jord, XXVI. Om mineraljordens och mullens rumsfyllande egenskaper. En teoretisk studie. (In Swedish). Swedish University of agricultural sciences, Uppsala.
- Andrade, A.R.S., da Paixao, F.J.R., de Azevedo, C.A.V., Neto, J.D., 2008. The soil available water estimate through the fractal theory. *Pesquisa Aplicada & Agrotecnologia*. V.1 N.1 Set.-Dez. 2008 ISSN 1983-6325.
- Arduino, E., Barberis, E., Boero, V., 1989. Iron-oxides and particle aggregation in B-horizons of some Italian soils. *Geoderma* 45, 319-329.
- Arrouays, D., Deslais, W., Bateau, V., 2001. The carbon content of topsoil and its geographical distribution in France. *Soil Use and Management* 17, 7-11.
- Arruda, F.B., Zullo, J., Oliveira, J.B., 1987. Parâmetros de solo para cálculo de água disponível com base na textura do solo. *Revista Brasileira De Ciencia Do Solo* 11, 11-15.
- Arshad, M., Ryan, J., Paeth, R.C., 1980. Influence of organic-matter, free iron and calcium-carbonate on soil particle-size distribution. *Agrochimica* 24, 470-477.
- Arya, L.M., Paris, J.F., 1981. A physicoempirical model to predict the soil-moisture characteristic from particle-size distribution and bulk-density data. *Soil Science Society of America Journal* 45, 1023-1030.
- Assouline, S., 2006. Modeling the relationship between soil bulk density and the water retention curve. *Vadose Zone Journal* 5, 554-563.
- Assouline, S., Tessier, D., Bruand, A., 1998. A conceptual model of the soil water retention curve. *Water Resources Research* 34, 223-231. doi:10.1029/97WR03039.

- Babalola, O., 1979. Spatial variability of soil water properties for a tropical soil of Nigeria. *Soil Science* 126, 269-279.
- Baert, G., 1995. Properties and chemical management aspects of soils on different parent rocks in the Lower Zaïre. PhD Thesis, Ghent University, Belgium.
- Baert, G., Embrechts, J., De Dapper, M., Mapaka, M., 1991a. Cartographie du Bas-Zaïre : Sols et évaluation des terres. Description et données analytiques des profils types du Bas-Zaïre. Universiteit Gent, Laboratorium voor Tropische Bodemkunde en Landevaluatie - ABOS.
- Baert, G., Embrechts, J., De Dapper, M., Mapaka, M., 1991b. Cartographie du Bas-Zaïre : Sols et évaluation des terres. Notice explicative : Feuilles de Luozi, de Tshela et Kai Mbaku, de Matadi, de Mbanza Ngungu, de Inkisi, de Kinshasa, de Banana. Description et données analytiques des profils types du Bas-Zaïre. Universiteit Gent, Laboratorium voor Tropische Bodemkunde en Landevaluatie - ABOS.
- Baert, G., Meyns, M., Van Ranst, E., Vanmechelen, L., 1993. Aménagement hydro-agricole de la Plaine de la Ruzizi, Zaïre.
- Baert, G., Van Ranst, E., 1998. Exchange properties of highly weathered soils of the Lower Congo. *Malaysian Journal of Soil Science* 2: 31-44.
- Baert, G., Van Ranst, E., Ngongo, M., Verdoodt, A., 2012. Soil survey in D.R. Congo – from 1935 until today. Paper presented at the meeting of the Section of Natural and Medical Sciences held on 27th March 2012. Brussels, Belgium.
- Baert, G., Van Ranst, E., Ngongo, M.L., Kasongo, E.L., Verdoodt, A., Mujinya, B.B., Mukalay, J., 2009. Guide des sols en R.D. Congo. Tome II : Description et données physico-chimiques de profils types. Imprimerie Salama-Don Bosco, Lubumbashi, RDC.
- Baert, G., Van Ranst, E., Vandenberghe, R., De Weirtdt, J., 1999. Estimation of Al-for-Fe substitution in Goethite by selective dissolution and Mössbauer spectroscopy in a weathering sequence on mafic rocks in the Lower Congo. *Malaysian Journal of Soil Science* 3, 11-27.
- Baeyens, J., 1938. Les sols de l'Afrique Centrale. Tome I : Le Bas Congo. Publication INEAC, Hors Série. INEAC, Bruxelles.
- Bagarello, V., Provenzano, G., Sgroi, A., 2009. Fitting particle size distribution models to data from Burundian soils for the BEST procedure and other purposes. *Biosystems Engineering* 104, 435-441.
- Bah, A.R., Kravchuk, O., Kirchhof, G., 2009. Fitting performance of particle-size distribution models on data derived by conventional and laser diffraction techniques. *Soil Science Society of America Journal* 73, 1101-1107.
- Baker, L., 2008. Development of class pedotransfer functions of soil water retention - A refinement. *Geoderma* 144, 225-230.
- Baker, L., Ellison, D., 2008a. Optimisation of pedotransfer functions using an artificial neural network ensemble method. *Geoderma* 144, 212-224.
- Baker, L., Ellison, D., 2008b. The wisdom of crowds - ensembles and modules in environmental modelling. *Geoderma* 147, 1-7.
- Baneejee, S.P., Pande, P., Pant, R.C., Subi, R.K., Bodala, K.C., 1981. Free iron oxide in relation to aggregation in lateritic soils of East Midnapur Forest Division, West Bengal. *Indian For.* 107, 24-29.
- Bannayan, M., Hoogenboom, G., 2009. Using pattern recognition for estimating cultivar coefficients of a crop simulation model. *Field Crops Research* 111, 290-302.
- Barral, M.T., Arias, M., Guerif, J., 1998. Effects of iron and organic matter on the porosity and structural stability of soil aggregates. *Soil & Tillage Research* 46, 261-272.

- Basile, A., Durso, G., 1997. Experimental corrections of simplified methods for predicting water retention curves in clay-loamy soils from particle-size determination. *Soil Technology* 10, 261-272.
- Batjes, N.H., 2008. Mapping soil carbon stocks of Central Africa using SOTER. *Geoderma* 146, 58-65.
- Bayat, H., Neyshaburi, M.R., Mohammadi, K., Nariman-Zadeh, N., Irannejad, M., Gregory, A.S., 2013. Combination of artificial neural networks and fractal theory to predict soil water retention curve. *Computers and Electronics in Agriculture* 92, 92-103.
- BEAU, 1987. Atlas du Bas-Zaïre. B.E.A.U., Kinshasa.
- Bell, M.A., Vankeulen, H., 1995. Soil pedotransfer functions for 4 mexican soils. *Soil Science Society of America Journal* 59, 865-871.
- Bellocchi, G., Rivington, M., Donatelli, M., Matthews, K., 2010. Validation of biophysical models: issues and methodologies. A review. *Agronomy for Sustainable Development* 30, 109-130.
- Benites, V.M., Machado, P., Fidalgo, E.C.C., Coelho, M.R., Madari, B.E., 2007. Pedotransfer functions for estimating soil bulk density from existing soil survey reports in Brazil. *Geoderma* 139, 90-97.
- Bernoux, M., Arrouays, D., Cerri, C., Volkoff, B., Jolivet, C., 1998. Bulk densities of Brazilian Amazon soils related to other soil properties. *Soil Science Society of America Journal* 62, 743-749.
- Bhavanarayana, M., Rao, T.V., Krishnamurthy, G.S.R., 1986. Statistical relationships of water retention and availability with soil matrix and charge properties. *International Agrophysics* 2, 135-144.
- Bigham, J.M., Golden, D.C., Bowen, L.H., Buol, S.W., Weed, S.B., 1978. Iron-oxide mineralogy of well-drained ultisols and oxisols .1. characterization of iron-oxides in soil clays by Mossbauer-spectroscopy, X-ray-diffractometry, and selected chemical techniques. *Soil Science Society of America Journal* 42, 816-825.
- Bird, N.R.A., Perrier, E., Rieu, M., 2000. The water retention function for a model of soil structure with pore and solid fractal distributions. *European Journal of Soil Science* 51, 55-63.
- Bittelli, M., Campbell, G.S., Flury, M., 1999. Characterization of particle-size distribution in soils with a fragmentation model. *Soil Science Society of America Journal* 63, 782-788.
- Blackmore, A.V., 1973. Aggregation of clay by the products of iron (III) hydrolysis. *Australian Journal of Soil Research* 11, 75-82.
- Borgesen, C.D., Schaap, M.G., 2005. Point and parameter pedotransfer functions for water retention predictions for Danish soils. *Geoderma* 127, 154-167.
- Borggaard, O.K., 1983. Iron-oxides in relation to aggregation of soil particles. *Acta Agriculturae Scandinavica* 33, 257-260.
- Bouma, J., 1989. Using soil survey data for quantitative land evaluation. *Advances in Soil Science* 9, 177-213.
- Bouma, J., Stoorvogel, J., van Alphen, B.J., Booltink, H.W.G., 1999. Pedology, precision agriculture, and the changing paradigm of agricultural research. *Soil Science Society of America Journal* 63, 1763-1768.
- Bouma, J., Van Lanen, H.A.J., 1987. Transfer functions and threshold values: from soil characteristics to land qualities. In: Beek, K.J., Burrough, P.A., McCormack, D.E. (Eds.), *Proceedings of the International Workshop on Quantified Land Evaluation Procedures*. Publication 6. ITC, Enschede, the Netherlands, pp. 106–110.
- Brady, N.C., Weil, R.R., 1999. *The nature and properties of soils*. Prentice Hall, Upper Saddle River, NJ.

- Braudeau, E., Frangi, J.P., Mohtar, R.H., 2004. Characterizing nonrigid aggregated soil-water medium using its shrinkage curve. *Soil Science Society of America Journal* 68, 359-370.
- Breiman, L., Friedman, J.H., Olshen, R.A., Stone, C.J., 1993. *Regression trees*. Chapman & Hall/CRC, Boca Raton, FL.
- Bronick, C.J., Lal, R., 2005. Soil structure and management: a review. *Geoderma* 124, 3-22.
- Brooks, R.H., Corey, A.T., 1964. Hydraulic properties of porous media. *Hydrology Paper* 3, Colorado State University, Fort Collins, Colorado, USA.
- Brouwers, M., 1987. Les sols de la Compagnie Sucrière de Kwilu Ngongo, Zaïre. Rapport de mission. CIRAD-ENSAM.
- Brown, J.D., Heuvelink, G.B.M., 2005. Assessing uncertainty propagation through physically based models of soil water flow and solute transport. In: Anderson, M.G. (Ed.), *Encyclopedia of hydrological sciences*. Wiley & Sons, Chichester, UK, pp. 1181-1196.
- Bruand, A., Duval, O., Cousin, I., 2004. Estimation des propriétés de rétention en eau des sols à partir de la base de données SOLHYDRO : une première proposition combinant le type d'horizon, sa texture, et sa densité apparente. *Etude et Gestion des Sols* 11, 323-332.
- Bruand, A., Fernandez, P.N., Duval, O., 2003. Use of class pedotransfer functions based on texture and bulk density of clods to generate water retention curves. *Soil Use and Management* 19, 232-242.
- Bruand, A., Perez-Fernandez, P., Duval, O., Quetin, P., Nicoullaud, B., Gaillard, H., Raison, L., Pessaud, J.F., Prud'Homme, L., 2002. Estimation des propriétés de rétention en eau des sols : utilisation de classe de pédotransfert après stratifications texturale et texturo-structurale. *Etude et gestion des sols* 9, 105-125.
- Brutsaert, W., 1966. Probability laws for the pore-size distributions. *Soil Science*, 101: 85-92.
- Buccigrossi, F., Caliandro, A., Rubino, P., Mastro, M.A., 2010. Testing some pedotransfer functions (PTFs) in Apulia region. Evaluation on the basis of soil particle size distribution and organic matter content for estimating field capacity and wilting point. *Italian Journal of Agronomy/Rivista di Agronomia* 4, 367-381.
- Buchan, G.D., 1989. Applicability of the simple lognormal model to particle-size distribution in soils. *Soil Science* 147, 155-161.
- Buchan, G.D., Grewal, K.S., Robson, A.B., 1993. Improved models of particle-size distribution - an illustration of model comparison techniques. *Soil Science Society of America Journal* 57, 901-908.
- Buishand, T.A., Brandsma, T., 2001. Multisite simulation of daily precipitation and temperature in the Rhine basin by nearest-neighbor resampling. *Water Resources Research* 37, 2761-2776.
- Bultot, F., 1950. Carte des régions climatiques du Congo belge établie d'après les critères de Köppen. *Communications N° 2 du Bureau Climat., INEAC. Collection in-4*.
- Cahen, L., Lepersonne, J., 1948. Notes sur la géomorphologie du Congo occidental. *Ann. Musée du Congo Belge, Bulletin N° 8*, 1: 7-95.
- Calhoun, F.G., Smeck, N.E., Slater, B.L., Bigham, J.M., Hall, G.F., 2001. Predicting bulk density of Ohio soils from morphology, genetic principles, and laboratory characterization data. *Soil Science Society of America Journal* 65, 811-819.
- Campbell, G.S., 1974. A simple method for determining unsaturated conductivity from moisture retention data. *Soil Science* 117, 311-314.
- Carsel, R.F., Parrish, R.S., 1988. Developing joint probability-distributions of soil-water retention characteristics. *Water Resources Research* 24, 755-769.

- Cassel, D.K., Lal, R., 1992. Soil physical properties of the tropics: common beliefs and management restraints. In: Lal, R., Sanches, P. (Eds.), *Myths and science of soils of the Tropics*, Special Publication N° 29. SSSA, Madison, WI, pp. 61–88.
- Cassel, D.K., Nielsen, D.R., 1986. Field capacity and available water capacity. In: Klute, A. (Ed.), *Methods of Soil Analysis. Part 1. Second Edition. Agronomy Monograph 9. ASA and SSSA*, Madison, WI, pp. 901–926.
- Cazemier, D.R., Lagacherie, P., Martin-Clouaire, R., 2001. A possibility theory approach for estimating available water capacity from imprecise information contained in soil databases. *Geoderma* 103, 113–132.
- Chakraborty, D., Mazumdar, S.P., Garg, R.N., Banerjee, S.; Santra, P., Singh, R., Tomar, R.K., 2011. Pedotransfer functions for predicting points on the moisture retention curve of Indian soils. *Indian Journal of Agricultural Sciences* 81, 1030–1036.
- Chesters, G., Attoe, O.J., Allen, O.N., 1957. Soil aggregation in relation to various soil constituents. *Soil Science Society of America Proceedings* 21, 272–277.
- Cihan, A., Perfect, E., Tyner, J.S., 2007. Water retention models for scale-variant and scale-invariant drainage of mass prefractal porous media. *Vadose Zone Journal* 6, 786–792.
- Cihan, A., Tyner, J.S., Perfect, E., 2009. Predicting relative permeability from water retention: A direct approach based on fractal geometry. *Water Resources Research* 45.
- Clapp, R.B., Hornberger, G.M., 1978. Empirical equations for some soil hydraulic-properties. *Water Resources Research* 14, 601–604.
- Clark, L.A., Pregibon, D., 1992. Tree-based models. In: Hastie, T.J. (Ed.), *Statistical Models in S*. Wadsworth, Pacific Grove, CA, pp. 377–419.
- Colombo, C., Torrent, J., 1991. Relationships between aggregation and iron-oxides in Terra Rossa soils from Southern Italy. *Catena* 18, 51–59.
- Compere, P., 1970. Carte des sols et de la végétation du Congo, du Rwanda et du Burundi. N° 25 : Bas-Congo, B, Végétation. INEAC.
- Cornelis, W.M., Corluy, J., Medina, H., Diaz, J., Hartmann, R., Van Meirvenne, M., Ruiz, M.E., 2006. Measuring and modelling the soil shrinkage characteristic curve. *Geoderma* 137, 179–191.
- Cornelis, W.M., Khlosi, M., Hartmann, R., Van Meirvenne, M., De Vos, B., 2005. Comparison of unimodal analytical expressions for the soil-water retention curve. *Soil Science Society of America Journal* 69, 1902–1911.
- Cornelis, W.M., Ronsyn, J., Van Meirvenne, M., Hartmann, R., 2001. Evaluation of pedotransfer functions for predicting the soil moisture retention curve. *Soil Science Society of America Journal* 65, 638–648.
- Correa, J.C., 1984. Características físico-hídricas dos solos latossolo amarelo, podzólico verm elho-amarelo e podzol hidromórfico do estado do Amazonas. *Pesquisa Agropecuária Brasileira* 19, 347–360.
- Cosby, B.J., Hornberger, G.M., Clapp, R.B., Ginn, T.R., 1984. A statistical exploration of the relationships of soil-moisture characteristics to the physical-properties of soils. *Water Resources Research* 20, 682–690.
- Cressie, N., 1993. *Statistics for spatial data. Revised Edition*. John Wiley & Sons, New York.
- Dasarathy, B.V. (Ed.), 1991. *Nearest neighbor (NN) Norms: NN pattern classification techniques*. IEEE Computer Society Press, Los Alamitos, CA.

- Dashtaki, S.G., Homaei, M., Khodaverdiloo, H., 2010. Derivation and validation of pedotransfer functions for estimating soil water retention curve using a variety of soil data. *Soil Use and Management* 26, 68-74.
- de Condappa, D., Galle, S., Dewandel, B., Haverkamp, R., 2008. Bimodal zone of the soil textural triangle: Common in tropical and subtropical regions. *Soil Science Society of America Journal* 72, 33-40.
- de Gennes, P.G., 1985. Partial filling of a fractal structure by a wetting fluid. In: Adler, D., Fritzsche, H., Ovshinsky, S.R. (Eds.), *Physics of disordered materials*. Plenum, New York, NY, pp. 227–241.
- De Vos, B., Van Meirvenne, M., Quataert, P., Deckers, J., Muys, B., 2005. Predictive quality of pedotransfer functions for estimating bulk density of forest soils. *Soil Science Society of America Journal* 69, 500-510.
- De Jong, R., Campbell, C.A., Nicholaichuk, W., 1983. Water-retention equations and their relationship to soil organic-matter and particle-size distribution for disturbed samples. *Canadian Journal of Soil Science* 63, 291-302.
- Demattê, J.L.I., 1988. Manejo de Solos ácidos dos Trópico úmidos – Região Amazônica. Fundação Cargill, Campinas, SP.
- Denef, K., Six, J., Merckx, R., Paustian, K., 2002. Short-term effects of biological and physical forces on aggregate formation in soils with different clay mineralogy. *Plant and Soil* 246, 185-200.
- Deng, H.L., Ye, M., Schaap, M.G., Khaleel, R., 2009. Quantification of uncertainty in pedotransfer function-based parameter estimation for unsaturated flow modeling. *Water Resources Research* 45, W04409, doi:10.1029/2008WR007477
- Denisoff, I., Devred, R., 1954. Carte des Sols et de la Végétation du Congo Belge et du Ruanda-Urundi. N° 2: Mvuazi (Bas Congo). Notice explicative. INEAC, Bruxelles.
- Devred, R., 1956. Les savanes herbeuses de la region de M’Vuazi (Bas-Congo). INEAC, Série Scient. N°37.
- Dijkerman, J.C., 1988. An Ustult-Aquult-Tropept catena in Sierra-Leone, West-Africa .2. Land qualities and land evaluation. *Geoderma* 42, 29-49.
- Dirksen, C., 1991. Unsaturated hydraulic conductivity. In: Smith, K.A., Mullins, C.E. (Eds.), *Soil Analysis: Physical Methods*. Marcel Dekker, Inc., New York, pp. 209–269.
- D’Urso, G., Basile, A., 1997. Physico-empirical approach for mapping soil hydraulic behaviour. *Hydrology and Earth System Sciences* 1, 915-923.
- Duvignaud, P., 1949. Les savanes du Bas-Congo. *Lejeunia, Mém.* N° 10.
- Efron, B., Tibshirani, R.J., 1993. An introduction to the Bootstrap. *Monographs on statistics and applied probability*. Chapman & Hall, New York, NY.
- Elshorbagy, A., Corzo, G., Srinivasulu, S., Solomatine, D.P., 2010a. Experimental investigation of the predictive capabilities of data driven modeling techniques in hydrology - Part 1: Concepts and methodology. *Hydrology and Earth System Sciences* 14, 1931-1941.
- Elshorbagy, A., Corzo, G., Srinivasulu, S., Solomatine, D.P., 2010b. Experimental investigation of the predictive capabilities of data driven modeling techniques in hydrology - Part 2: Application. *Hydrology and Earth System Sciences* 14, 1943-1961.
- Espino, A., Mallants, D., Vanclooster, M., Feyen, J., 1996. Cautionary notes on the use of pedotransfer functions for estimating soil hydraulic properties. *Agricultural Water Management* 29, 235-253.
- Eswaran, H., Vandenberg, E., Reich, P., 1993. Organic-carbon in soils of the world. *Soil Science Society of America Journal* 57, 192-194.
- FAO, 1974. Management properties of Ferralsols. *FAO Soils Bull.* 23, FAO, Rome.

- FAO, 2002. Crops and drops. Making the best use of water for agriculture. Rome, Italy.
- FAO-UNESCO, 1974. Soil map of the world. Volume I: legend. UNESCO, Paris.
- FAO-UNESCO-ISRIC., 1990. Soil map of the world: revised legend. World Soil Resources Report 60, Reprinted, Rome. 138 pp.
- Farkas, C., Rajkai, K., Kertész, M., Bakacsi, Z., van Meirvenne, M., 2008. Spatial variability of soil hydro-physical properties: A case study in Herceghalom, Hungary. In: Krasilnikov, P., Carré, F. Montanarella, L. (Eds.), Soil geography and geostatistics. Concepts and applications. European Commission. Joint Research Centre. Institute for Environment and Sustainability, Ispra, Italy.
- Farlow, S.J., 1984. The GMDH algorithm. In: Farlow, S.J. (Ed.), Self-organizing methods in modeling: GMDH-type algorithms. M. Dekker, New York, pp. 1–24.
- Feder, J., 1988. Fractals. Plenum Press, New York.
- Federer, C.A., 1983. Nitrogen mineralization and nitrification - depth variation in 4 New-England forest soils. Soil Science Society of America Journal 47, 1008-1014.
- Fernandez-Illescas, C.P., Porporato, A., Laio, F., Rodriguez-Iturbe, I., 2001. The ecohydrological role of soil texture in a water-limited ecosystem. Water Resources Research 37, 2863-2872.
- Ferrer-Julia, M., Estrela-Monreal, T., Sanchez del Corral-Jimenez, A., Garcia-Melendez, E., 2004. Constructing a saturated hydraulic conductivity map of Spain using pedotransfer functions and spatial prediction. Geoderma 123, 257-277.
- Finke, P.A., 1995. Soil inventorization: from data collection to providing information (in Dutch). In: Buurman, P., Sevink, J. (Eds.), From soil map to information system. Wageningen Press, Wageningen, the Netherlands, pp. 111–125.
- Finke, P.A., Wösten, J.H.M., Jansen, M.J.W., 1996. Effects of uncertainty in major input variables on simulated functional soil behaviour. Hydrological Processes 10, 661-669.
- Fredlund, M.D., Fredlund, D.G., Wilson, G.W., 2000. An equation to represent grain-size distribution. Canadian Geotechnical Journal 37, 817-827.
- Friedl, J., Schwertmann, U., 1996. Aluminium influence on iron oxides .18. The effect of Al substitution and crystal size magnetic hyperfine fields of natural goethites. Clay Minerals 31, 455-464.
- Friedman, J.H., 1991. Multivariate adaptive regression splines. Annals of Statistics 19, 1-67.
- Gee, G.W., Bauder, J.W., 1986. Particle-size analysis. In: Klute, A. (Ed.), Methods of Soil Analysis. Part 1. Second Edition. Agronomy Monograph 9. ASA and SSSA, Madison, WI, pp. 383–411.
- Gessler, P.E., 1996. Statistical soil-landscape modelling for environmental management. PhD Thesis, Australian National University, Canberra, Australia.
- Gessler, P.E., Moore, I.D., McKenzie, N.J., Ryan, P.J., 1995. Soil-landscape modeling and spatial prediction of soil attributes. International Journal of Geographical Information Systems 9, 421-432.
- Ghanbarian-Alavijeh, B., Liaghat, A., Huang, G.H., Van Genuchten, M.Th., 2010. Estimation of the van Genuchten soil water retention properties from soil textural data. Pedosphere 20, 456-465.
- Ghanbarian-Alavijeh, B., Millan, H., 2010. Point pedotransfer functions for estimating soil water retention curve. International Agrophysics 24, 243-251.
- Gharahi Ghehi, N., Nemes, A., Verdoodt, A., Van Ranst, E., Cornelis, W.M., Boeckx, P., 2012. Nonparametric techniques for predicting soil bulk density of tropical rainforest topsoils in Rwanda. Soil Science Society of America Journal 76, 1172-1183.
- Gijsman, A.J., Jagtap, S.S., Jones, J.W., 2002. Wading through a swamp of complete confusion: how to choose a method for estimating soil water retention parameters for crop models. European Journal of Agronomy 18, 75-105.

- Gimenez, D., Rawls, W.J., Pachepsky, Y., Watt, J.P.C., 2001. Prediction of a pore distribution factor from soil textural and mechanical parameters. *Soil Science* 166, 79-88.
- Givi, J., Prasher, S.O., Patel, R.M., 2004. Evaluation of pedotransfer functions in predicting the soil water contents at field capacity and wilting point. *Agricultural Water Management* 70, 83-96.
- Goldberg, S., Kapoor, B.S., Rhoades, J.D., 1990. Effect of aluminum and iron-oxides and organic-matter on flocculation and dispersion of arid zone soils. *Soil Science* 150, 588-593.
- Goovaerts, P., 1997. *Geostatistics for natural resource evaluation*. Oxford University Press, New York.
- Gosselink, J.G., Hatton, R., Hopkinson, C.S., 1984. Relationship of organic-carbon and mineral-content to bulk-density in Louisiana marsh soils. *Soil Science* 137, 177-180.
- Grossman, R.B., Reinsch, T.G., 2002. Bulk density and linear extensibility. In: Dane, J.H., Topp, G.C. (Eds.), *Methods of soil analysis. Part 4. Physical methods*. SSSA Madison Wisconsin, USA, pp. 201–228.
- Guber, A.K., Pachepsky, Y.A., 2010. *Multimodeling with Pedotransfer Functions. Documentation and user manual for PTF Calculator (CalcPTF) Version 3.0*. Environmental Microbial and Food Safety Laboratory, Beltsville Agricultural Research Center, USDA-ARS.
- Guber, A.K., Pachepsky, Y.A., van Genuchten, M.Th., Rawls, W.J., Šimůnek, J., Jacques, D., Nicholson, T.J., Cady, R.E., 2006. Field-scale water flow simulations using ensembles of pedotransfer functions for soil water retention. *Vadose Zone Journal* 5, 234-247.
- Guber, A.K., Pachepsky, Y.A., van Genuchten, M.Th., Šimůnek, J., Jacques, D., Nemes, A., Nicholson, T.J., Cady, R.E., 2009. Multimodel Simulation of Water Flow in a Field Soil Using Pedotransfer Functions. *Vadose Zone Journal* 8, 1-10.
- Gupta, S.C., Larson, W.E., 1979. Estimating soil-water retention characteristics from particle-size distribution, organic-matter percent, and bulk-density. *Water Resources Research* 15, 1633-1635.
- Haghverdi, A., Cornelis, W.M., Ghahraman, B., 2012. A pseudo-continuous neural network approach for developing water retention pedotransfer functions with limited data. *Journal of Hydrology* 442, 46-54.
- Handreck, K.A., 1983. Particle-size and the physical-properties of growing media for containers. *Communications in Soil Science and Plant Analysis* 14, 209-222.
- Hansen, V.E., Israelsen, O.W., Stringham, G.E., 1980. *Irrigation principles and practices*. Fourth Edition. Wiley, New York.
- Harrison, A.F., Bockock, K.L., 1981. Estimation of soil bulk-density from loss-on-ignition values. *Journal of Applied Ecology* 18, 919-927.
- Hastie, T., Tibshirani, R., Friedman, J., 2009. *The elements of statistical learning: Prediction, inference and data mining*. Second Edition. Springer-Verlag, New York.
- Hastie, T.J., Pregibon, D., 1992. Generalized linear models. In: Chambers, J.M., Hastie, T.J. (Eds.), *Statistical Models in S*. Wadsworth and Brooks, California, USA, pp. 195–248.
- Hastie, T.J., Tibshirani, R., 1990. *Generalized Additive Models*. Chapman & Hall, London.
- Hastie, T.J., Tibshirani, R., Friedman, J., 2001. *The elements of statistical learning: data mining, inference and prediction*. Springer Series in Statistics. Springer-Verlag, New York.
- Haverkamp, R., Parlange, J.-Y., 1986. Predicting the water retention curve from a particle size distribution: 1. Sandy soils without organic matter. *Soil Science* 142, 325-339.
- Haykin, S., 1994. *Neural Networks, a comprehensive foundation*. First Edition. Macmillan College Publishing Company, New York, NY.
- Hecht-Nielsen, R., 1990. *Neurocomputing*. Addison-Wesley, Reading, MA.

- Herbst, M., Diekkrüger, B., 2002. The influence of the spatial structure of soil properties on water balance modeling in a microscale catchment. *Physics and Chemistry of the Earth, Part B* 27, 701–710.
- Hillel, D., 1980. *Fundamentals of soil physics*. Academic Press, New York.
- Hillel, D., 1998. *Environmental Soil Physics*. Academic Press, San Diego.
- Hodnett, M.G., Tomasella, J., 2002. Marked differences between van Genuchten soil water-retention parameters for temperate and tropical soils: a new water-retention pedo-transfer function developed for tropical soils. *Geoderma* 108, 155-180.
- Hollis, J.M., Hannam, J., Bellamy, P.H., 2012. Empirically-derived pedotransfer functions for predicting bulk density in European soils. *European Journal of Soil Science* 63, 96-109.
- Honeysett, J.L., Ratkowsky, D.A., 1989. The use of ignition loss to estimate bulk-density of forest soils. *Journal of Soil Science* 40, 299-308.
- Hong, S.Y., Minasny, B., Han, K.H., Kim, Y., Lee, K. 2013. Predicting and mapping soil available water capacity in Korea. *PeerJ* 1:e71. doi 10.7717/peerj.71.
- Hopmans, J.W., 1987. A comparison of various methods to scale soil hydraulic-properties. *Journal of Hydrology* 93, 241-256.
- Hopmans, J.W., Šimůnek, J., Romano, N., Durner, W., 2002. Water retention and storage: Inverse methods. In: Dane, J.H., Topp, G.C. (Eds.), *Methods of soil analysis. Part 4. Physical methods*. SSSA Book Series N° 5. SSSA, Madison, WI, pp. 963–1004.
- Howard, P.J.A., Loveland, P.J., Bradley, R.I., Dry, F.T., Howard, D.M., Howard, D.C., 1995. The carbon content of soil and its geographical-distribution in Great-Britain. *Soil Use and Management* 11, 9-15.
- Huang, G.H., Zhang, R.D., 2005. Evaluation of soil water retention curve with the pore-solid fractal model. *Geoderma* 127, 52-61.
- Huang, S., Yang, Y., Wang, Y., 2003. A critical look at procedures for validating growth and yield models. In: Amaro, A., Reed, D., Soares, P. (Eds.), *Modelling Forest Systems*. CABI Publishing, Cambridge, MA, pp. 271–293.
- Huntington, T.G., Johnson, C.E., Johnson, A.H., Siccama, T.G., Ryan, D.F., 1989. Carbon, organic-matter, and bulk-density relationships in a forested Spodosol. *Soil Science* 148, 380-386.
- Hwang, S.I., 2004. Effect of texture on the performance of soil particle-size distribution models. *Geoderma* 123, 363-371.
- Hwang, S.I., Choi, S.I., 2006. Use of a lognormal distribution model for estimating soil water retention curves from particle-size distribution data. *Journal of Hydrology* 323, 325-334.
- Hwang, S.I., Lee, K.P., Lee, D.S., Powers, S.E., 2002. Models for estimating soil particle-size distributions. *Soil Science Society of America Journal* 66, 1143-1150.
- Hwang, S.I., Powers, S.E., 2003. Lognormal distribution model for estimating soil water retention curves for sandy soils. *Soil Science* 168, 156-166.
- Hwang, S.I., Yun, E.Y., Ro, H.M., 2011. Estimation of soil water retention function based on asymmetry between particle- and pore-size distributions. *European Journal of Soil Science* 62, 195-205.
- IFAGRARIA, 1975. *Projet de développement intégré de la vallée de la Luaa. Rapport final*. Kinshasa.
- Igwe, C.A., Akamigbo, F.O.R., Mbagwu, J.S.C., 2002. Soil moisture retention characteristics in relation to erodibility and texture of some soils of Southeastern Nigeria. *East African Agricultural and Forestry Journal* 68, 17-21.
- Imam, B., Sorooshian, S., Mayr, T., Schaap, M.G., Wösten, J.H.M., Scholes, R.J., 1999. Comparison of pedotransfer functions to compute water holding capacity using the Van Genuchten model in

- inorganic soils. Report to IGBP-DIS soil data tasks. IGBP-DIS Working Paper 22. IGBP-DIS, Toulouse, Cedex, France.
- Isaaks, E., Srivastava, R., 1989. An introduction to Applied Geostatistics. Oxford University Press, New York.
- ISSS Working Group, 1998. World reference base for soil resources: Introduction. First Edition. International Society of Soil Science, International Soil Reference and Information Centre, and Food and Agriculture Organization of the United Nations, Acco, Leuven, the Netherlands.
- IUSS Working Group WRB, 2006. World reference base for soil resources 2006. Second Edition. World Soil Resources Reports 103. FAO, Rome.
- Jagtap, S.S., Lall, U., Jones, J.W., Gijsman, A.J., Ritchie, J.T., 2004. Dynamic nearest-neighbor method for estimating soil water parameters. Transactions of the ASAE 47, 1437-1444.
- Jaky, J., 1944. Soil mechanics. (In Hungarian). Egyetemi Nyomda, Budapest.
- Jalabert, S.S.M., Martin, M.P., Renaud, J.P., Boulonne, L., Jolivet, C., Montanarella, L., Arrouays, D., 2010. Estimating forest soil bulk density using boosted regression modelling. Soil Use and Management 26, 516-528.
- Jamagne, M., Betremieux, R., Begon, J.C., Mori, A., 1977. Quelques données sur la variabilité dans le milieu naturel de la réserve en eau des sols. agro - INRA 127, 627-641.
- Jamagne, M., 1963. Contribution à l'étude des sols au Congo oriental. PEDOLOGIE XIII(2), 271-414.
- Jauhiainen, M., 2004. Relationships of particle size distribution curve, soil water retention curve and unsaturated hydraulic conductivity and their implications on water balance of forested and agricultural hillslopes. PhD Thesis, Helsinki University of Technology, Finland.
- Jeffrey, D.W. 1970. A note on the use of ignition loss as a means for the approximate estimation of soil bulk density. The Journal of Ecology, 58, 297-299.
- Jin, Z., Dong, Y.S., Qi, Y.C., Liu, W.G., An, Z.S., 2011. Characterizing variations in soil particle-size distribution along a grass-desert shrub transition in the Ordos Plateau of Inner Mongolia, China. Land Degradation & Development doi:10.1002/ldr.1112.
- Jing'an, C., Guojiang, W., 1999. Sediment particle size distribution and its environmental significance in Lake Erhai, Yunnan Province. Chinese Journal of Geochemistry 4(18), 314-320.
- Johnson, N.L., Kotz, S., 1970. Distributions in statistics. Continuous univariate distributions. Volume 1. Wiley, New York.
- Jost, C., 1977. Les sols du Bas-Zaïre. B.E.A.U., Kinshasa.
- Journel, A.G., Huijbregts, C.H.J., 1978. Mining Geostatistics. New York, Academic Press.
- Juo, A.S.R., Franzluebbers, K., 2003. Tropical soils: properties and management for sustainable agriculture. Oxford University Press, Oxford, England.
- Jury, W.A., Horton, R., 2004. Soil Physics. Sixth Edition. John Wiley & Sons, Inc., Hoboken, NJ.
- Kar, G., Kumar, A., Singh, R., 2009. Spatial distribution of soil hydro-physical properties and morphometric analysis of a rainfed watershed as a tool for sustainable land use planning. Agricultural Water Management 96, 1449-1459.
- Kar, G., Singh, R., Verma, H.N., 2004. Spatial variability studies of soil hydro-physical properties using GIS for sustainable crop planning of a watershed of eastern India and its testing in a rainfed rice area. Australian Journal of Soil Research 42, 369-379.
- Kaur, R., Kumar, S., Gurung, H.P., 2002. A pedo-transfer function (PTF) for estimating soil bulk density from basic soil data and its comparison with existing PTFs. Australian Journal of Soil Research 40, 847-857.

- Kern, J.S., 1995. Evaluation of soil water retention models based on basic soil physical properties. *Soil Science Society of America Journal* 59, 1134-1141.
- Khlosi, M., 2003. Physical-conceptual models to predict the soil moisture retention curve. MSc Thesis, Ghent University, Belgium.
- Koekkoek, E.J.W., Booltink, H., 1999. Neural network models to predict soil water retention. *European Journal of Soil Science* 50, 489-495.
- Köhn, M., 1929. Korngrößenanalyse vermittle Pipettanalyse. *Tonindustrie-Zeitung* 53, 729-731.
- Kolev, B., Rousseva, S., Dimitrov, D., 1996. Derivation of soil water capacity parameters from standard soil texture information for Bulgarian soils. *Ecological Modelling* 84, 315-319.
- Koy, K.R., 2005. Impact of planted acacia forest on the chemical fertility of sandy soils of the bateke plateau (D.R. Congo). MSc Thesis, Ghent University, Belgium.
- Koza, J., 1992. *Genetic Programming: On the Programming of Computers by Natural Selection*. MIT Press, Cambridge, MA.
- Kravchenko, A., Zhang, R.D., 1998. Estimating the soil water retention from particle-size distributions: A fractal approach. *Soil Science* 163, 171-179.
- Krishna Murti, G.S.R., Huang, P.M., 1987. Influence of constituents on the stability of mechanical separates of soils representing major taxonomic orders. *Appl. Clay Sci.* 2, 299-308.
- Lal, R., 1978. Physical-properties and moisture retention characteristics of some Nigerian soils. *Geoderma* 21, 209-223.
- Lal, R., 1981. Physical properties. In: Greenland, D.J. (Ed.), *Characterization of soils in relation to their classification and management for crop production: Examples from some areas in the humid tropics*. Clarendon Press, Oxford, pp. 135-148.
- Lamorski, K., Pachepsky, Y., Slawinski, C., Walczak, R.T., 2008. Using support vector machines to develop pedotransfer functions for water retention of soils in Poland. *Soil Science Society of America Journal* 72, 1243-1247.
- Lane, P.W., 2002. Generalized linear models in soil science. *European Journal of Soil Science* 53, 241-251.
- Leenaars, J.G.B., 2013. Africa soil profiles database, Version 1.1. A compilation of georeferenced and standardised legacy soil profile data for Sub-Saharan Africa (with dataset). ISRIC Report 2013/03. Africa Soil Information Service (AfSIS) project and ISRIC — World Soil Information, Wageningen, the Netherlands (160 pp.; 16 fig.; 10 tab.; 27 ref).
- Leonavičiūtė, N., 2000. Predicting soil bulk and particle densities by pedotransfer functions from existing soil data in Lithuania. *Geografijos metraštis* 33, 317-330.
- Li, J., Heap, A.D., 2011. A review of comparative studies of spatial interpolation methods in environmental sciences: Performance and impact factors. *Ecological Informatics* 6, 228-241.
- Li, Y., Wardlaw, N.C., 1986. Mechanisms of nonwetting phase trapping during imbibition at slow rates. *Journal of Colloid and Interface Science* 109, 473-486.
- Lima, J.E.F.W., Silva, E.M., 2007. Seleção de modelos para o traçado de curvas granulométricas de sedimentos em suspensão em rios. *Revista Brasileira de Engenharia Agrícola e Ambiental* 11, 101-107.
- Lin, H.S., 2003. Hydropedology: bridging disciplines, scales, and data. *Vadose Zone Journal* 2, 1-11.
- Loosvelt, L., Pauwels, V.R.N., Cornelis, W.M., De Lannoy, G.J.M., Verhoest, N.E.C., 2011. Impact of soil hydraulic parameter uncertainty on soil moisture modeling. *Water Resources Research* 47.
- MacLean, A.H., Yager, T.U., 1972. Available water capacities of Zambian soils in relation to pressure plate measurements and particle size analysis. *Soil Science* 113, 23-29.

- Mahieu, J., 1961. Reboisement des savanes et enrichissement des groupements de colonisation au Mayumbe. INEAC.
- Makkeasorn, A., Chang, N.B., Beaman, M., Wyatt, C., Slater, C., 2006. Soil moisture estimation in a semiarid watershed using RADARSAT-1 satellite imagery and genetic programming. *Water Resources Research* 42, doi:10.1029/2005WR004033.
- Mandelbrot, B.B., 1983. *The fractal geometry of nature*. W.H. Freeman, San Francisco, CA.
- Manrique, L.A., Jones, C.A., 1991. Bulk-density of soils in relation to soil physical and chemical-properties. *Soil Science Society of America Journal* 55, 476-481.
- Marcelino, V.M.M., 1995. Characteristics and genesis of sandy soils of Lower Zaire. PhD Thesis, Ghent University, Belgium.
- Martin, M.P., Lo Seen, D., Boulonne, L., Jolivet, C., Nair, K.M., Bourgeon, G., Arrouays, D., 2009. Optimizing pedotransfer functions for estimating soil bulk density using boosted regression trees. *Soil Science Society of America Journal* 73, 485-493.
- Mason, G., Morrow, N.R., 1991. Capillary behavior of a perfectly wetting liquid in irregular triangular tubes. *Journal of Colloid and Interface Science* 141, 262-274.
- Matthews, G.P., Watts, C.W., Powlson, D.S., Price, J.C., Whalley, W.R., 2008. Wetting of agricultural soil measured by a simplified capillary rise technique. *European Journal of Soil Science* 59, 817-823.
- Maudoux, E., 1954. La régénération naturelle dans les forêts remaniées du Mayumbe. *Bulletin Agric. du Congo Belge* 45, 403-421.
- McBratney, A.B., Minasny, B., Cattle, S.R., Vervoort, R.W., 2002. From pedotransfer functions to soil inference systems. *Geoderma* 109, 41-73.
- McBratney, A.B., Minasny, B., Tranter, G., 2011. Necessary meta-data for pedotransfer functions. *Geoderma* 160, 627-629.
- McBratney, A.B., Santos, M.L.M., Minasny, B., 2003. On digital soil mapping. *Geoderma* 117, 3-52.
- McKenzie, N.J., Austin, M.P., 1993. A quantitative Australian approach to medium and small-scale surveys based on soil stratigraphy and environmental correlation. *Geoderma* 57, 329-355.
- Mdemu, M.V., Mulengera, M.K., 2002. Using pedotransfer functions (PTFs) to estimate soil water retention characteristics (SWRCs) in the tropics for sustainable soil water management: Tanzania Case Study. 12th ISCO Conference, Beijing.
- Medina, H., Tarawally, M., del Valle, A., Ruiz, M.E., 2002. Estimating soil water retention curve in Rhodic Ferralsols from basic soil data. *Geoderma* 108, 277-285.
- Mehra, O.P., Jackson, M.L., 1960. Iron oxide removal from soils and clays by a dithionite-citrate system buffered with sodium bicarbonate. *Clays Clay Miner.* 7, 317-327.
- Mekaru, T., Uehara, G., 1972. Anion adsorption in ferruginous tropical soils. *Soil Science Society of America Proceedings* 36, 296-300.
- Millan, H., Aguilar, M., Dominguez, J., Cespedes, L., Velasco, E., Gonzalez, M., 2006. A note on the physics of soil water retention through fractal parameters. *Fractals-Complex Geometry Patterns and Scaling in Nature and Society* 14, 143-148.
- Millan, H., Gonzalez-Posada, M., 2005. Modelling soil water retention scaling. Comparison of a classical fractal model with a piecewise approach. *Geoderma* 125, 25-38.
- Millington, R., Quirk, J.P., 1961. Permeability of porous solids. *Transactions of the Faraday Society* 57, 1200-1207.
- Milly, P.C.D., Shmakin, A.B., 2002. Global modeling of land water and energy balances. Part I: the land dynamics (LaD) model. *Journal of Hydrometeorology* 3, 283-299.

- Minasny, B., Hartemink, A.E., 2011. Predicting soil properties in the tropics. *Earth-Science Reviews* 106, 52-62.
- Minasny, B., McBratney, A.B., 2002. The neuro-m method for fitting neural network parametric pedotransfer functions. *Soil Science Society of America Journal* 66, 352-361.
- Minasny, B., McBratney, A.B., Bristow, K.L., 1999. Comparison of different approaches to the development of pedotransfer functions for water-retention curves. *Geoderma* 93, 225-253.
- Minasny, B., McBratney, A.B., 2002b. Neuropack. Neural Network package for fitting pedotransfer functions. Technical Note. v 1.0. From the Australian Centre for Precision Agriculture.
- Moeys, J., Larsbo, M., Bergstrom, L., Brown, C.D., Coquet, Y., Jarvis, N.J., 2012. Functional test of pedotransfer functions to predict water flow and solute transport with the dual-permeability model MACRO. *Hydrology and Earth System Sciences* 16, 2069-2083.
- Moeys, J., Shangguan, W., 2010. Soil texture: Functions for soil texture plot, classification and transformation. <http://cran.r-project.org/web/packages/soiltexture/>. Accessed 21 May 2012.
- Molina, N.C., Caceres, M.R., Pietroboni, A.M., 2001. Factors affecting aggregate stability and water dispersible clay of recently cultivated semiarid soils of Argentina. *Arid Land Research and Management* 15, 77-87.
- Moreels, E., De Neve, S., Hofman, G., Van Meirvenne, M., 2003. Simulating nitrate leaching in bare fallow soils. A model comparison. *Nutrient Cycling in Agroecosystems* 67, 137-144.
- Mpulila, T.S., Cornelis, W.M., Van Ranst, E., 2013. Effect of clay mineralogy on water retention of soils in the Eastern and Northern Tanzania (submitted).
- Mualem, Y., 1976. A new model for predicting the hydraulic conductivity of unsaturated porous media. *Water Resources Research* 12, 513-522.
- Mucherino, A., Papajorgji, P., Pardalos, P.M., 2009. Data mining in agriculture. Springer, New York.
- Nebel, A.L.C., Timm, L.C., Cornelis, W., Gabriels, D., Reichardt, K., Aquino, L.S., Pauletto, E.A., Reinert, D.J., 2010. Pedotransfer functions related to spatial variability of water retention attributes for lowland soils. *Revista Brasileira De Ciencia Do Solo* 34, 669-680.
- Neimark, A.V., 1989. Multiscale percolation systems. *Soviet Physics –JETP* 69: 786-791.
- Nemes, A., Pachepsky, Y.A., Timlin, D.J., 2011. Toward improving global estimates of field soil water capacity. *Soil Science Society of America Journal* 75, 807-812.
- Nemes, A., Quebedeaux, B., Timlin, D.J., 2010. Ensemble approach to provide uncertainty estimates of soil bulk density. *Soil Science Society of America Journal* 74, 1938-1945.
- Nemes, A., Rawls, W.J., 2004. Soil texture and particle-size distribution as input to estimate soil hydraulic properties. In: Pachepsky, Y.A., Rawls, W.J. (Eds.), *Development of pedotransfer functions in soil hydrology*. Elsevier, Amsterdam, pp. 47–70.
- Nemes, A., Rawls, W.J., Pachepsky, Y.A., 2005. Influence of organic matter on the estimation of saturated hydraulic conductivity. *Soil Science Society of America Journal* 69, 1330-1337.
- Nemes, A., Rawls, W.J., Pachepsky, Y.A., 2006a. Use of the nonparametric nearest neighbor approach to estimate soil hydraulic properties. *Soil Science Society of America Journal* 70, 327-336.
- Nemes, A., Rawls, W.J., Pachepsky, Y.A., van Genuchten, M.Th., 2006b. Sensitivity analysis of the nonparametric nearest neighbor technique to estimate soil water retention. *Vadose Zone Journal* 5, 1222-1235.
- Nemes, A., Roberts, R.T., Rawls, W.J., Pachepsky, Y.A., van Genuchten, M.Th., 2008. Software to estimate 33 and 1500 kPa soil water retention using the non-parametric k-nearest neighbor technique. *Environmental Modelling & Software* 23, 254-255.

- Nemes, A., Schaap, M.G., Leij, F.J., Wösten, J.H.M., 2001. Description of the unsaturated soil hydraulic database UNSODA version 2.0. *Journal of Hydrology* 251, 151-162.
- Nemes, A., Schaap, M.G., Wösten, J.H.M., 2003. Functional evaluation of pedotransfer functions derived from different scales of data collection. *Soil Science Society of America Journal* 67, 1093-1102.
- Nemes, A., Wösten, J.H.M., Lilly, A., Voshaar, J.H.O., 1999. Evaluation of different procedures to interpolate particle-size distributions to achieve compatibility within soil databases. *Geoderma* 90, 187-202.
- Nicolai, H., 1961. Luozi. Géographie régionale d'un pays du Bas-Congo. CEMUBAC. Extrait de Mém. Académie Royale des Sciences d'Outre-Mer.
- Noble, W.S., 2006. What is a support vector machine? *Nature Biotechnology* 24, 1565-1567.
- Oades, J.M., Waters, A.G., 1991. Aggregate hierarchy in soils. *Australian Journal of Soil Research* 29, 815-828.
- Obalum, S.E., Obi, M.E., 2012. Moisture characteristics and their point pedotransfer functions for coarse-textured tropical soils differing in structural degradation status. *Hydrological Processes*. (accepted).
- Oliveira, L.B., Ribeiro, M.R., Jacomine, P.K.T., Rodrigues, J.J.V., Marques, F.A., 2002. Funções de pedotransferência para predição de umidade retida a potencias específicos em solos do estado de Pernambuco. *Revista Brasileira De Ciencia Do Solo* 26, 315-323.
- Ottoni Filho, T.B., Ottoni, M.V., 2010. A variation of the field capacity (FC) definition and a FC database for Brazilian soils. In: Gilkes, R.J., Prakongkep, N. (Eds.), *Soil solutions for a changing world: 19th World Congress of Soil Science*. 1–6 August 2010. IUSS, Brisbane, QLD, Australia.
- Pachepsky, Y., Rawls, W., Gimenez, D., Watt, J.P.C., 1998. Use of soil penetration resistance and group method of data handling to improve soil water retention estimates. *Soil & Tillage Research* 49, 117-126.
- Pachepsky, Y.A., Crawford, J.W., Rawls, W.J. (Eds.), 2000. *Fractals in Soil Science*. Developments in soil science 27: 1-295.
- Pachepsky, Y.A., Gimenez, D., Lilly, A., Nemes, A., 2008. Soil structure and soil hydrologic functioning: a hydropedological framework. *International Hydropedology Conference*, Penn State, University Park, PA, July 28-31, 2008.
- Pachepsky, Y.A., Polubesova, T.A., Hajnos, M., Sokolowska, Z., Jozefaciuk, G., 1995. Fractal parameters of pore surface-area as influenced by simulated soil degradation. *Soil Science Society of America Journal* 59, 68-75.
- Pachepsky, Y.A., Radcliffe, D.E., Selim, H.M., 2003. *Scaling methods in soil physics*. CRC Press, Boca Raton, FL.
- Pachepsky, Y.A., Rawls, W.J., 2003. Soil structure and pedotransfer functions. *European Journal of Soil Science* 54, 443-451.
- Pachepsky, Y.A., Rawls, W.J., Lin, H.S., 2006. Hydropedology and pedotransfer functions. *Geoderma* 131, 308-316.
- Pachepsky, Y.A., Schaap, M.G., 2004. Data mining and exploration techniques. In: Pachepsky, Y.A., Rawls, W.J. (Eds.), *Development of pedotransfer functions in soil hydrology*. Elsevier, Amsterdam, pp. 21–32.
- Pachepsky, Y.A., Timlin, D., Varallyay, G., 1996. Artificial neural networks to estimate soil water retention from easily measurable data. *Soil Science Society of America Journal* 60, 727-733.

- Parasuraman, K., Elshorbagy, A., 2008. Toward improving the reliability of hydrologic prediction: Model structure uncertainty and its quantification using ensemble-based genetic programming framework. *Water Resources Research* 44, W12406, doi:10.1029/2007WR006451.
- Parasuraman, K., Elshorbagy, A., Carey, S.K., 2007a. Modelling the dynamics of the evapotranspiration process using genetic programming. *Hydrological Sciences Journal-Journal Des Sciences Hydrologiques* 52, 563-578.
- Parasuraman, K., Elshorbagy, A., Si, B.C., 2006. Estimating saturated hydraulic conductivity in spatially variable fields using neural network ensembles. *Soil Science Society of America Journal* 70, 1851-1859.
- Parasuraman, K., Elshorbagy, A., Si, B.C., 2007b. Estimating saturated hydraulic conductivity using genetic programming. *Soil Science Society of America Journal* 71, 1676-1684.
- Patil, N.G., Pal, D.K., Mandal, C., Mandal, D.K., 2012. Soil water retention characteristics of vertisols and pedotransfer functions based on nearest neighbor and neural networks approaches to estimate AWC. *Journal of Irrigation and Drainage Engineering-ASCE* 138, 177-184.
- Patil, N.G., Rajput, G.S., 2009. Evaluation of water retention functions and computer program "Rosetta" in predicting soil water characteristics of seasonally impounded shrink-swell soils. *Journal of Irrigation and Drainage Engineering-ASCE* 135, 286-294.
- Patil, N.G., Rajput, G.S., Nema, R.K., Singh, R.B., 2009. Calibration and evaluation of pedotransfer functions to estimate available water capacity of seasonally impounded shrink-swell soils of Central India. *Archives of Agronomy and Soil Science*.
- Perfect, E., 1999. Estimating soil mass fractal dimensions from water retention curves. *Geoderma* 88, 221-231.
- Perfect, E., Kay, B.D., 1995. Applications of fractals in soil and tillage research: A review. *Soil & Tillage Research* 36, 1-20.
- Perfect, E., McLaughlin, N.B., Kay, B.D., Topp, G.C., 1998. Comment on "An improved fractal equation for the soil water retention curve" by E. Perfect et al. - Reply. *Water Resources Research* 34, 933-935.
- Perkins, K., Nimmo, J., 2009. High-quality unsaturated zone hydraulic property data for hydrologic applications. *Water Resources Research* 45, W07417, doi:10.1029/2008WR007497
- Perrier, E., Bird, N., Rieu, M., 1999. Generalizing the fractal model of soil structure: the pore-solid fractal approach. *Geoderma* 88, 137-164.
- Perrier, E., Rieu, M., Sposito, G., deMarsily, G., 1996. Models of the water retention curve for soils with a fractal pore size distribution. *Water Resources Research* 32, 3025-3031.
- Pidgeon, J.D., 1972. Measurement and prediction of available water capacity of ferrallitic soils in Uganda. *Journal of Soil Science* 23, 431-441.
- Piedallu, C., Gegout, J.C., Bruand, A., Seynave, I., 2011. Mapping soil water holding capacity over large areas to predict potential production of forest stands. *Geoderma* 160, 355-366.
- Pinstrup-Andersen, P., Pandhya-Lorch, R., Rosengrant, M.W., 1999. *World Food Prospects: Critical Issues for the Early Twenty-First Century*. International Food Policy Research Institute, Washington, DC.
- PNUD/UNOPS, 1998a. *Monographie de la province du Bas-Congo*. Ministère de l'Agriculture et de l'Elevage, du Plan, de l'Education Nationale et de l'Environnement, Conservation de la nature, Forêts et Pêche, Kinshasa, R.D. Congo.

- PNUD/UNOPS, 1998b. Monographie de la ville de Kinshasa. Ministère de l'Agriculture et de l'Elevage, du Plan, de l'Education Nationale et de l'Environnement, Conservation de la nature, Forêts et Pêche, Kinshasa, R.D. Congo.
- Puckett, W.E., Dane, J.H., Hajek, B.F., 1985. Physical and mineralogical data to determine soil hydraulic-properties. *Soil Science Society of America Journal* 49, 831-836.
- Rajkai, K., Kabos, S., van Genuchten, M.Th., 2004. Estimating the water retention curve from soil properties: comparison of linear, nonlinear and concomitant variable methods. *Soil & Tillage Research* 79, 145-152.
- Rao, K.S., Narasimha Rao, P.V., Mohan, B.K., Venketachalam, P., 1988. Relation between water retention characteristics of soil and their physical properties. *Journal of Soil and Water Conservation in India* 32, 52-69.
- Rawls, W.J., 1983. Estimating soil bulk-density from particle-size analysis and organic-matter content. *Soil Science* 135, 123-125.
- Rawls, W.J., Brakensiek, D.L., 1982. Estimating soil-water retention from soil properties. *Journal of the Irrigation and Drainage Division-ASCE* 108, 166-171.
- Rawls, W.J., Brakensiek, D.L., 1985. Prediction of soil water properties for hydrological modeling. In: Jones, E.B., Wards, T.J. (Eds.), *Watershed management in the eighties*. Proc. of Symp. 30th April 30 – 1st May. Sponsored by Comm. on Watershed Management, I & D Division, ASCE. ASCE Convention, Denver, Colorado, USA, pp. 293–299.
- Rawls, W.J., Brakensiek, D.L., Saxton, K.E., 1982. Estimation of soil-water properties. *Transactions of the ASAE* 25, 1316-1320.
- Rawls, W.J., Brakensiek, D.L., Soni, B., 1983. Agricultural management effects on soil-water processes .1. soil-water retention and Green and Ampt infiltration parameters. *Transactions of the ASAE* 26, 1747-1752.
- Rawls, W.J., Gish, T.J., Brakensiek, D.L. 1991. Estimating soil water retention from soil physical properties and characteristics. *Advances in Soil Science* 16, 213-234.
- Rawls, W.J., Pachepsky, Y.A., Ritchie, J.C., Sobecki, T.M., Bloodworth, H., 2003. Effect of soil organic carbon on soil water retention. *Geoderma* 116, 61-76.
- Reichardt, K., 1988. Capacidade de campo. *Revista Brasileira De Ciencia Do Solo* 12, 211-216.
- Reichert, J.M., Albuquerque, J.A., Kaiser, D.R., Reinert, D.J., Urach, F.L., Carlesso, R., 2009. Estimation of water retention and availability in soils of Rio Grande do Sul. *Revista Brasileira De Ciencia Do Solo* 33, 1547-1560.
- Rieu, M., Sposito, G., 1991a. Fractal fragmentation, soil porosity, and soil-water properties .1. Theory. *Soil Science Society of America Journal* 55, 1231-1238.
- Rieu, M., Sposito, G., 1991b. Fractal fragmentation, soil porosity, and soil-water properties .2. Applications. *Soil Science Society of America Journal* 55, 1239-1244.
- Romano, N., Santini, A., 1997. Effectiveness of using pedo-transfer functions to quantify the spatial variability of soil water retention characteristics. *Journal of Hydrology* 202, 137-157.
- Rouault, Y., Assouline, S., 1998. A probabilistic approach towards modeling the relationships between particle and pore size distributions: the multicomponent packed sphere case. *Powder Technology* 96, 33-41.
- Ruehlmann, J., Korschens, M., 2009. Calculating the effect of soil organic matter concentration on soil bulk density. *Soil Science Society of America Journal* 73, 876-885.
- Saada, A., Siffert, B., Papirer, E., 1995. Comparison of the hydrophilicity/hydrophobicity of illites and kaolinites. *Journal of Colloid and Interface Science* 174, 185-190.

- Saikia, U.S., Singh, A.K., 2003. Development and validation of pedotransfer functions for water retention, saturated hydraulic conductivity and aggregate stability of soils of Banha watershed. *Journal of the Indian Society of Soil Science* 51, 484-488.
- Sanchez, P.A., 2002. Ecology - Soil fertility and hunger in Africa. *Science* 295, 2019-2020.
- Santos, W.J.R. dos, Curi, N., Silva, S.H.G., ARAÚJO, E.F. de, Marques, J.J., 2013. Pedotransfer functions for water retention in different soil classes from the center-southern Rio Grande do Sul state. *Ciência e Agrotecnologia, Lavras*, V.37, n.1.
- Santra, P., Chopra, U.K., Chakraborty, D., 2008. Spatial variability of soil properties and its application in predicting surface map of hydraulic parameters in an agricultural farm. *Current Science* 7, 937-945.
- Santra, P., Das, B.S., 2008. Pedotransfer functions for soil hydraulic properties developed from a hilly watershed of Eastern India. *Geoderma* 146, 439-448.
- Saxton, K.E., Rawls, W.J., 2006. Soil water characteristic estimates by texture and organic matter for hydrologic solutions. *Soil Science Society of America Journal* 70, 1569-1578.
- Saxton, K.E., Rawls, W.J., Romberger, J.S., Papendick, R.I., 1986. Estimating generalized soil-water characteristics from texture. *Soil Science Society of America Journal* 50, 1031-1036.
- Saxton, K.E., Willey, P.H., 2006. The SPAW model for agricultural field and pond hydrologic simulation. In: Singh, V.P., Frevert, D.K. (Eds.), *Watershed Models*. CRC Press, Boca Raton, FL, pp. 401-435.
- Schaap, M.G., 2004. Accuracy and uncertainty in PTF predictions. In: Pachepsky, Y.A., Rawls, W.J. (Eds.), *Development of Pedotransfer functions in soil hydrology*. Elsevier, Amsterdam, pp. 33-43.
- Schaap, M.G., 2005. Models for indirect estimation of soil hydraulic properties. In: Anderson, M. (Ed.), *Encyclopedia of Hydrological Sciences*. John Wiley & Sons, Ltd.
- Schaap, M.G., Leij, F.J., 1998. Using neural networks to predict soil water retention and soil hydraulic conductivity. *Soil & Tillage Research* 47, 37-42.
- Schaap, M.G., Leij, F.J., van Genuchten, M.Th., 1999. A bootstrap-neural network approach to predict soil hydraulic parameters. In: van Genuchten, M.Th., Leij, F.J., Wu, L. (Eds.), *Proc. Int. Workshop, Characterization and measurements of the hydraulic properties of unsaturated porous media*. University of California, Riverside, CA, pp. 1237-1250.
- Schaap, M.G., Leij, F.J., van Genuchten, M.Th., 2001. ROSETTA: a computer program for estimating soil hydraulic parameters with hierarchical pedotransfer functions. *Journal of Hydrology* 251, 163-176.
- Scheinost, A.C., Sinowski, W., Auerswald, K., 1997. Regionalization of soil water retention curves in a highly variable soilscape .1. Developing a new pedotransfer function. *Geoderma* 78, 129-143.
- Schulze, D.G., Schwertmann, U., 1984. The influence of aluminum on iron-oxides .10. Properties of Al-substituted goethites. *Clay Minerals* 19, 521-539.
- Schwartz, D., Namri, M., 2002. Mapping the total organic carbon in the soils of the Congo. *Global and Planetary Change* 33, 77-93.
- Schwartz, R.C., Evett, S.R., 2002. Estimating hydraulic properties of a fine-textured soil using a disc infiltrometer. *Soil Science Society of America Journal* 66, 1409-1423.
- Schwertmann, U., 1988a. Some properties of soil and synthetic iron oxides. In: Stucki, J.W Goodman, B.A., Schwertmann, U. (Eds.), *Iron in Soils and Clay Minerals*. NATO ASI Series, v. 217, pp. 203-250.

- Schwertmann, U., 1988b. Occurrence and formation of iron oxides in various environments. In: Stucki, J.W. et al. (Eds.), *Iron in soils and clay minerals*, Nato Adv. Study Inst., Ser. C, Vol. 217: pp. 267-308.
- Schwertmann, U., Fitzpatrick, R. W., Taylor, R. M., Lewis, D. G., 1979. The influence of aluminum on iron oxides. Part II. Preparation and properties of Al-substituted hematites. *Clay and Clay Minerals* 2, 105-112.
- Schwertmann, U., Kampf, N., 1985. Properties of goethite and hematite in kaolinitic soils of Southern and Central Brazil. *Soil Science* 139, 344-350.
- Schwertmann, U., Latham, M., 1986. Properties of iron-oxides in some new Caledonian Oxisols. *Geoderma* 39, 105-123.
- Schwertmann, U., Taylor, R.M., 1989. Iron oxides. In: Dixon, J.B., Weed, S.B. (Eds.), *Minerals in soil environment*. Second Edition. SSSA, Madison, Wisconsin, pp. 379-438.
- Selle, B., Muttill, N., 2011. Testing the structure of a hydrological model using genetic programming. *Journal of Hydrology* 397, 1-9.
- Shangguan, W., 2012. An investigation of soil particle-size distribution models for the conversion of soil texture classification from ISSS and Katschinski's to FAO/USDA System. Paper presented at PEDOFRACT VII Workshop on Scaling in Particulate and Porous Media: Modeling and Use in Predictions. A Coruña, Spain.
- Shangguan, W., Dai, Y.J., Liu, B.Y., Ye, A.Z., Yuan, H., 2012. A soil particle size distribution dataset for regional land and climate modelling in China. *Geoderma* 171, 85-91.
- Sharma, M.L., Uehara, G., 1968. Influence of soil structure on water relations in low Humic Latosols .I. Water retention. *Soil Science Society of America Proceedings* 32, 765-770.
- Sharma, S.K., Mohanty, B.P., Zhu, J.T., 2006. Including topography and vegetation attributes for developing pedotransfer functions. *Soil Science Society of America Journal* 70, 1430-1440.
- Shein, E.V., Arkhangel'skaya, T.A., 2006. Pedotransfer functions: state of the art, problems, and outlooks. *Eurasian Soil Science* 10, 1089-1099.
- Shepherd, K.D., Walsh, M.G., 2002. Development of reflectance spectral libraries for characterization of soil properties. *Soil Science Society of America Journal* 66, 988-998.
- Silva, E.M., Lima, J.E.F.W., Rodrigues, L.N., Azevedo, J.A., 2004. Comparação de modelos matemáticos para o traçado de curvas granulométricas. *Pesquisa Agropecuária Brasileira* 39, 363-370.
- Šimůnek, J., Hopmans, J.W., 2002. Parameter optimization and nonlinear fitting. In: Dane, J.H., Topp, G.C. (Eds.), *Methods of Soil Analysis. Part 4. Physical Methods*. Third Edition. SSSA, Madison, WI, pp. 139-157.
- Singh, A.K., 2000. Use of pedotransfer functions in crop growth simulation. *Journal of Water Management* 8, 18-21.
- Six, J., Elliott, E.T., Paustian, K., 2000. Soil structure and soil organic matter: II. A normalized stability index and the effect of mineralogy. *Soil Science Society of America Journal* 64, 1042-1049.
- Skaggs, T.H., Arya, L.M., Shouse, P.J., Mohanty, B.P., 2001. Estimating particle-size distribution from limited soil texture data. *Soil Science Society of America Journal* 65, 1038-1044.
- Smith, P., Smith, J.U., Powlson, D.S., McGill, W.B., Arah, J.R.M., Chertov, O.G., Coleman, K., Franko, U., Frolking, S., Jenkinson, D.S., Jensen, L.S., Kelly, R.H., Klein-Gunnewiek, H., Komarov, A.S., Li, C., Molina, J.A.E., Mueller, T., Parton, W.J., Thornley, J.H.M., Whitmore, A.P., 1997. A comparison of the performance of nine soil organic matter models using datasets from seven long-term experiments. *Geoderma* 81, 153-225.

- Smith, R.E., Diekkruger, B., 1996. Effective soil water characteristics and ensemble soil water profiles in heterogeneous soils. *Water Resources Research* 32, 1993-2002.
- SOCINCO, 1959a. Etude des plateaux et vallées dans la zone de Thysville. Rapport d'étude, Kinshasa.
- SOCINCO, 1959b. Etude de Kasi. Rapport d'étude, Kinshasa.
- SOCINCO, 1960. Etude de la vallée de la Luala. Rapport d'étude, Vol. 1, 2 et 3. Kinshasa.
- SOGREAH, 1959. Mise en valeur du Stanley Pool. Rapport d'étude, Kinshasa.
- SOGREAH-SOCINCO, 1964. Prospection systématique des sols à l'Ouest de la rivière Kwango. Rapport d'étude. C.E.E., Kinshasa.
- Soil Survey Staff, 1975. Soil Taxonomy. Soil Conservation Service, USDA, Agriculture Handbook 436. US Govt. Printing Office, Washington, D.C.
- Soil Survey Staff, 1990. Keys to Soil Taxonomy. SMSS Tech. Monogr. 19, 4th ed. Virginia Polytechnic Inst. and State Univ., Blacksburg. Soil Conserv. Serv., USDA, Washington, D.C.
- Soil Survey Staff, 1992. Keys to Soil Taxonomy. SMSS Tech. Monogr. 19, Fifth Edition. Virginia Polytechnic Inst. and State Univ., Blacksburg. Soil Conserv. Serv., USDA, Washington, D.C.
- Soil Survey Staff., 1995. Soil Characterization and Profile Description Data. CD-ROM. Soil Survey Laboratory, Natural Resources Conservation Service, USDA, Lincoln, NE.
- Sorensen, D.C., 1982. Newton's method with a model trust region modification", *SIAM Journal on Numerical Analysis* 19, 409-426.
- Steduto, P., Hsiao, T.C., Raes, D., Fereres, E., 2009. AquaCrop—The FAO crop model to simulate yield response to water. I. Concepts and underlying principles. *Agronomy Journal* 101, 426-437.
- Steenstra, B., 1988. De geomorfologie van Beneden-Zaire. Académie Royale des Sciences d'Outre-Mer. Symposium 'De maritieme toegang tot Zaire'.
- Stewart, V.I., Adams, W.A., Abdulla, H.H., 1970. Quantitative pedological studies on soils derived from Silurian mudstones .2. Relationship between stone content and apparent density of fine earth. *Journal of Soil Science* 21, 248-255.
- Stoops, G., 1966. Bijdrage tot de kennis van de bodenvorming in Beneden-Congo (Léopoldville). PhD Thesis, Ghent University, Belgium.
- Su, Y.Z., Zhao, H.L., Zhao, W.Z., Zhang, T.H., 2004. Fractal features of soil particle size distribution and the implication for indicating desertification. *Geoderma* 122, 43-49.
- Suuster, E., Ritz, C., Roostalu, H., Reintam, E., Kolli, R., Astover, A., 2011. Soil bulk density pedotransfer functions of the humus horizon in arable soils. *Geoderma* 163, 74-82.
- Sys, C., 1972. Caractérisation morphologique et physico-chimique des profils types de l'Afrique centrale. Publication INEAC, Hors Série. INEAC, Bruxelles.
- Sys, C., Van Ranst, E., Debaveye, J., 1991. Land Evaluation. Part I. General Administration for Development Cooperation, Brussels, Belgium.
- Sys, C., Van Wambeke, A., Frankart, R., Gilson, P., Jongen, P., Pecrot, A., Berce, J.-M., Jamagne, M., 1961. La cartographie des sols au Congo : ses principes et ses méthodes. Publication INEAC, Série technique, N°66. INEAC, Bruxelles.
- Tamminen, P., Starr, M., 1994. Bulk density of forested mineral soils. *Silva Fennica* 28, 53-60.
- Taylor, S.A., Ashcroft, G.M., 1972. Physical edaphology. Freeman and Co. San Francisco, CA.
- Tempel, P., Batjes, N.H., van Engelen, V.W.P., 1996. IGBP-DIS soil data set for pedotransfer function development. Working Paper 96/05. ISRIC, Wageningen, the Netherlands.
- Tietje, O., Hennings, V., 1993. Evaluation of pedotransfer functions for estimating the water-retention curve. *Zeitschrift Fur Pflanzenernahrung Und Bodenkunde* 156, 447-455.

- Tietje, O., Tapkenhinrichs, M., 1993. Evaluation of pedotransfer functions. *Soil Science Society of America Journal* 57, 1088-1095.
- Tomasella, J., Hodnett, M.G., 1998. Estimating soil water retention characteristics from limited data in Brazilian Amazonia. *Soil Science* 163, 190-202.
- Tomasella, J., Hodnett, M.G., 2004. Pedotransfer functions for tropical soils. In: Pachepsky, Y.A., Rawls, W.J. (Eds.), *Development of pedotransfer functions in soil hydrology*. Elsevier, Amsterdam, pp. 415-429.
- Tomasella, J., Hodnett, M.G., Rossato, L., 2000. Pedotransfer functions for the estimation of soil water retention in Brazilian soils. *Soil Science Society of America Journal* 64, 327-338.
- Tomasella, J., Pachepsky, Y., Crestana, S., Rawls, W.J., 2003. Comparison of two techniques to develop pedotransfer functions for water retention. *Soil Science Society of America Journal* 67, 1085-1092.
- Tombu, C., Huart, A., 2012. Bas-Congo. Weyrich Édition, Neufchâteau, Belgique.
- Torrent, J., 1994. Iron oxides in mediterranean soils: properties and influence on soil behavior. 15th World Congress of Soil Science, Acapulco, v.8a, pp. 2-14.
- Tranter, G., Minasny, B., McBratney, A.B., Murphy, B., McKenzie, N.J., Grundy, M., Brough, D., 2007. Building and testing conceptual and empirical models for predicting soil bulk density. *Soil Use and Management* 23, 437-443.
- Tuller, M., Or, D., Dudley, L.M., 1999. Adsorption and capillary condensation in porous media: Liquid retention and interfacial configurations in angular pores. *Water Resources Research* 35, 1949-1964.
- Twarakavi, N.K.C., Šimůnek, J., Schaap, M.G., 2009. Development of pedotransfer functions for estimation of soil hydraulic parameters using support vector machines. *Soil Science Society of America Journal* 73, 1443-1452.
- Tyler, S.W., Wheatcraft, S.W., 1990. Fractal processes in soil-water retention. *Water Resources Research* 26, 1047-1054.
- Tyler, S.W., Wheatcraft, S.W., 1992. Fractal scaling of soil particle-size distributions - analysis and limitations. *Soil Science Society of America Journal* 56, 362-369.
- Ungaro, F., Calzolari, C., Busoni, E., 2005. Development of pedotransfer functions using a group method of data handling for the soil of the Pianura Padano-Veneta region of North Italy: water retention properties. *Geoderma* 124, 293-317.
- van den Berg, M., 1996. Available water capacity in strongly weathered soils of South East and Southern Brazil. *Solo Suelo* 96, XIII Congresso Latinoamericano de Ciencia do Solo Aguas de Lindoia, SP-Brazil, August 4-8, 1996 (Proceedings on CD-ROM).
- van den Berg, M., Klamt, E., vanReeuwijk, L.P., Sombroek, W.G., 1997. Pedotransfer functions for the estimation of moisture retention characteristics of Ferralsols and related soils. *Geoderma* 78, 161-180.
- Van Genuchten, M.Th., 1980. A closed-form equation for predicting the hydraulic conductivity of unsaturated soils. *Soil Science Society of America Journal* 44, 892-898.
- van Genuchten, M.Th., Leij, F.J., Yates, S.R., 1991. The RETC code for quantifying the hydraulic functions of unsaturated soils. USDA, US Salinity Laboratory, Riverside, CA. United States Environmental Protection Agency, document EPAR600r2-91r065.
- Van Meirvenne, M., Denaeghel, J., Rajkai, K., Kertész, M., Hofman, G., 1995. Spatial extension of a point water balance model. In: Schoute, J.F.Th. et al. (Eds.), *Scenario Studies for the Rural Environment*. Kluwer, the Netherlands, pp. 293-297.

- Van Meirvenne, M., Scheldeman, K., Baert, G., Hofman, G., 1994. Quantification of soil textural fractions of bas-zaire using soil map polygons and or point observations. *Geoderma* 62, 69-82.
- Van Ranst, E., 1991. Soils of the tropics and the subtropics: geography, classification, properties and management. Lecture notes. Ghent University, Belgium.
- Van Ranst, E., 1995. Rational soil management in the humid tropics. *Mededelingen Der Zittingen Koninklijke Academie voor Overzeese Wetenschappen* 40, 209-233.
- Van Ranst, E., 1997. Soil Mineralogy. Lecture notes. Ghent University, Belgium.
- Van Ranst, E., Baert, G., Ngongo, M., Mafuka, P. (Eds.), 2010. Carte pédologique de la province du Bas-Congo et ville de Kinshasa, échelle 1:500.000. R.D. Congo. UGent, Hogent, UNILU & UNIKIN (ISBN 9789491114625).
- Van Ranst, E., Shamshuddin, J., Baert, G., Dzwowa, P.K., 1998. Charge characteristics in relation to free iron and organic matter of soils from Bambouto mountains, western Cameroon. *European Journal of Soil Science* 49, 243-252.
- Van Ranst, E., Verloo, M., Demeyer, A., Pauwels, J.M., 1999. Manual for the soil chemistry and fertility laboratory. analytical methods for soils and plants. Equipment and management of consumables. International Training Centre for Post-Graduate Soil Scientists. Universiteit Gent, Gent, Belgium.
- Vandam, J.C., Stricker, J.N.M., Droogers, P., 1994. Inverse method to determine soil hydraulic functions from multistep outflow experiments. *Soil Science Society of America Journal* 58, 647-652.
- Vapnik, V., 1995. The nature of statistical learning theory. Springer, New York.
- Vapnik, V., 1998. Statistical learning theory. John Wiley & Sons, New York.
- Vaz, C.M.P., Iossi, M.D., Naime, J.D., Macedo, A., Reichert, J.M., Reinert, D.J., Cooper, M., 2005. Validation of the Arya and Paris water retention model for Brazilian soils. *Soil Science Society of America Journal* 69, 577-583.
- Vazquez, E.V., Ferreiro, J.P., Miranda, J.G.V., Gonzalez, A.P., 2008. Multifractal analysis of pore size distributions as affected by simulated rainfall. *Vadose Zone Journal* 7, 500-511.
- Velayutham, M., Raj, D., 1971. Interrelationship between soil separates and properties of the soils of Tamil Nadu. *Journal of the Indian Society of Soil Science* 19, 353-361.
- Vereecken, H., 1995. Estimating the unsaturated hydraulic conductivity from theoretical-models using simple soil properties. *Geoderma* 65, 81-92.
- Vereecken, H., Herbst, M., 2004 Statistical regression. . In: Pachepsky, Y.A., Rawls, W.J. (Eds.), *Development of pedotransfer functions in soil hydrology*. Elsevier, Amsterdam, pp. 415–429.
- Vereecken, H., Maes, J., Feyen, J., Darius, P., 1989. Estimating the soil-moisture retention characteristic from texture, bulk-density, and carbon content. *Soil Science* 148, 389-403.
- Vereecken, H., Weynants, M., Javaux, M., Pachepsky, Y., Schaap, M.G., van Genuchten, M.Th., 2010. Using Pedotransfer Functions to Estimate the van Genuchten-Mualem Soil Hydraulic Properties: A Review. *Vadose Zone Journal* 9, 795-820.
- Von Uexküll, H.R., Mutert, E., 1995. Global extent, development and economic impact of acid soils. In: Date, R.A., Grundon, N.J., Rayment, G.E., Probert, M.E. (Eds.), *Plant-soil interactions at low pH*. Kluwer Academic, pp. 749–752.
- Walkley, A. Black, I.A., 1934. An examination of the Degtjareff method for determining soil organic matter, and a proposed modification of the chromic soil titration method. *Soil Science* 37, 29-38.
- Wang, D., Fu, B.J., Zhao, W.W., Hu, H.F., Wang, Y.F., 2008. Multifractal characteristics of soil particle size distribution under different land-use types on the Loess Plateau, China. *Catena* 72, 29-36.

- Weynants, M., Vereecken, H., Javaux, M., 2009. Revisiting Vereecken pedotransfer functions: Introducing a closed-form hydraulic model. *Vadose Zone Journal* 8, 86-95.
- Whitmore, A.P., 1991. A method for assessing the goodness of computer-simulation of soil processes. *Journal of Soil Science* 42, 289-299.
- Wikipedia, 2013a. <https://fr.wikipedia.org/wiki/Bas-Congo> (consulted on 1st June 2013).
- Wikipedia, 2013b. <https://fr.wikipedia.org/wiki/Kinshasa> (consulted on 1st June 2013).
- Williams, J., Prebble, R.E., Williams, W.T., Hignett, C.T., 1983. The influence of texture, structure and clay mineralogy on the soil-moisture characteristic. *Australian Journal of Soil Research* 21, 15-32.
- Wösten, J.H.M., Bouma, J., Stoffelsen, G.H., 1985. Use of soil survey data for regional soil-water simulation-models. *Soil Science Society of America Journal* 49, 1238-1244.
- Wösten, J.H.M., Finke, P.A., Jansen, M.J.W., 1995. Comparison of class and continuous pedotransfer functions to generate soil hydraulic characteristics. *Geoderma* 66, 227-237.
- Wösten, J.H.M., Lilly, A., Nemes, A., Le Bas, C., 1999. Development and use of a database of hydraulic properties of European soils. *Geoderma* 90, 169-185.
- Wösten, J.H.M., Pachepsky, Y.A., Rawls, W.J., 2001. Pedotransfer functions: bridging the gap between available basic soil data and missing soil hydraulic characteristics. *Journal of Hydrology* 251, 123-150.
- Wösten, J.H.M., Schuren, C., Bouma, J., Stein, A., 1990. Functional sensitivity analysis of 4 methods to generate soil hydraulic functions. *Soil Science Society of America Journal* 54, 832-836.
- Wösten, J.H.M., Verzandvoort, S.J.E., Leenaars, J.G.B., Hoogland, T., Wesseling, J.G., 2013. Soil hydraulic information for river basin studies in semi-arid regions. *Geoderma* 195, 79-86.
- Wu, Q., Borkovec, M., Sticher, H., 1993. On particle-size distributions in soils. *Soil Science Society of America Journal* 57, 883-890.
- Yang, J., Greenwood, D.J., Rowell, D.L., Wadsworth, G.A., Burns, I.G., 2000. Statistical methods for evaluating a crop nitrogen simulation model, N_ABLE. *Agricultural Systems* 64, 37-53.
- Yerima, B.P.K., Van Ranst, E., 2005. *Introduction to Soil Science: Soils of the Tropics*. Trafford Publishing, Canada & U.K.
- Zacharias, S., Wessolek, G., 2007. Excluding organic matter content from pedotransfer predictors of soil water retention. *Soil Science Society of America Journal* 71, 43-50.
- Zhao, P., Shao, M.A., Horton, R., 2011. Performance of soil particle-size distribution models for describing deposited soils adjacent to constructed dams in the China Loess Plateau. *Acta Geophysica* 59, 124-138.
- Zhuang, J., Jin, Y., Miyazaki, T., 2001. Estimating water retention characteristic from soil particle-size distribution using a nonsimilar media concept. *Soil Science* 166, 308-321.

Curriculum vitae

1. Personal information

- ❖ First name: Yves-Dady
- ❖ Family name: BOTULA Manyala
- ❖ Birthplace: Kinshasa
- ❖ Birthdate: , 14th August 1981
- ❖ Country of origin: The Democratic Republic of Congo
- ❖ E-mail: ydbotula@yahoo.fr; ydbotula@gmail.com; YvesDady.BotulaManyala@UGent.be

2. Graduate studies

- ❖ February 2004: BSc. in Agronomy, Department of Natural Resources Management (Fauna and Flora), University of Kinshasa, D.R. Congo.

3. Post-graduate studies

- ❖ September 2007: MSc. in Physical Land Resources, Soil Science, Ghent University, Belgium.
- ❖ September 2008 up to now: Doctorate in Applied Biological Sciences: Land and Forest Management, Department of Soil Management, Faculty of Bioscience Engineering, Ghent University, Belgium.

4. Research and professional experience

- ❖ From March 2004 to August 2005: Research Assistant under the supervision of Prof. Dr. ir. Paul Mafuka Mbe-Mpie at the Department of Natural Resources Management, Faculty of Agronomy, University of Kinshasa, D.R. Congo.
- ❖ From September 2005 to September 2007: Post-graduate research experience in the International Course Program Master in Physical Land Resources, Option: Soil Science at Ghent University, Belgium.
- ❖ From October 2007 to March 2008: Research Assistant under the supervision of Prof. Dr. ir. Paul Mafuka Mbe-Mpie at the Department of Natural Resources Management, Faculty of Agronomy, University of Kinshasa, D.R. Congo.
- ❖ From March 2008 to July 2008: Landscape Assistant Consultant in the Salonga Project, Worldwide Fund for Nature (WWF), D.R. Congo.
- ❖ From September 2008 up to now: Doctoral student and research staff at the Department of Soil Management, Faculty of Bio-Science Engineering at Ghent University, Belgium.

5. Doctoral training program related to research

- ❖ Seminars in transferable skills on:
 - Research Methodology
 - Sampling for Survey and Monitoring of Natural Resources
 - AquaCrop Workshop FAO crop water productivity model
 - Advanced Academic English: Conference Skills
 - Advanced Academic English: Writing Skills
- ❖ Post-graduate courses on:
 - Applied Linear Statistics
 - Multivariate Data Analysis
 - Merging measurements and modelling in Soil Physics
- ❖ Intensive hands-on training on the k-Nearest Neighbour (k-NN) algorithm in College Park, Maryland (USA) under the supervision of Dr. Attila Nemes, primary developer of a variant of the k-NN algorithm for prediction of soil water retention.

6. Research results

6.1. Articles A-1

- ❖ Botula, Y.D., Cornelis, W.M., Baert, G., Van Ranst, E., 2012. Evaluation of pedotransfer functions for predicting water retention of soils in the Lower Congo (D.R. Congo). *Agricultural Water Management Journal* 111, 1-10.
- ❖ Botula, Y.-D., Cornelis, W.M., Baert, G., Mafuka, P., Van Ranst, E., 2013a. Particle size distribution models for soils of the humid tropics. *Journal of Soils and Sediments* 13, 686-698.
- ❖ Botula, Y.-D., Nemes, A., Mafuka, P., Van Ranst, E., Cornelis, W.M., 2013. Prediction of water content of soils from the humid tropics by the non-parametric k-nearest neighbor approach. *Vadose Zone Journal* doi:10.2136/vzj2012.0123.

- ❖ Botula, Y.-D., Van Ranst, E., Cornelis, W.M., 2013. Pedotransfer functions to predict water retention of soils of the humid tropics: a review (article under revision in Revista Brasileira de Ciencia do Solo).

6.2. Articles A-1 in preparation

- ❖ Hierarchical pedotransfer functions to predict bulk density of Lower Congo soils.
- ❖ Hierarchical Pedotransfer functions to predict water retention of soils in Lower Congo.
- ❖ Performance of three pore-solid fractal approaches for soil water retention of highly weathered soils of the humid tropics.
- ❖ Evaluation of pedotransfer functions for mapping of soil hydrophysical properties in Lower Congo.

6.3. Conferences

2009: Performance of pedotransfer functions to predict water retention of soils in Lower Congo (D.R. Congo) at the 3rd International Conference Geologica Belgica, Ghent, 14-15 September 2009 (poster session).

2010: Bridging hydrology and pedology to predict physical and hydraulic properties of soils in Lower Congo at the Thematic Day of the Soil Science Society of Belgium: 'History and future of soil research in Congo, Rwanda and Burundi', Brussels, 1st December 2010 (oral presentation).

2011: Bridging hydrology and pedology: An answer to the lack of soil hydraulic properties data in the Lower Congo at the Young Soil Scientist Day of the Soil Science Society of Belgium, Brussels, 23rd February 2011 (oral presentation).

2012: k-Nearest Neighbor approach for predicting water retention of soils in the Lower Congo (D.R. Congo) at the 17th PhD Symposium on Applied Biological Sciences in Leuven, 10th February 2012 and at the Young Soil Scientist Day of the Soil Science Society of Belgium, Brussels, 22nd February 2012 (poster session).

2012: Fitting performance of particle size distribution models for soils of the humid tropics at 4th Eurosoil International Congress, Bari, 2-6 July (poster session).

2012: Use of the nonparametric nearest neighbor approach to estimate water content of soils in the humid tropics at 4th Eurosoil International Congress, Bari, 2-6 July (poster session).

2012: Erosion control through the vetiver system in urban settings in Central Africa at the 67th International Annual conference of the Soil water conservation Society, Fort Worth, Texas, 22-25 July 2012 (oral presentation).

2012: Opportunities of conservation agriculture-based resources conserving practices to build resilience against drought and soil degradation in Northern Ethiopia at the 67th International Annual conference of the Soil water conservation Society, Fort Worth, Texas, 22-25 July 2012 (oral presentation).

2013: Predicting soil hydraulic properties in the humid tropics by a non-parametric approach at the 10th Biennial Meeting of Commission 1.5 Pedometrics Division 1 of the International Union of Soil Science (IUSS), Nairobi, 26-31 August 2013 (oral presentation)..

6.4. Tutorship

2010: Tutor of the MSc. thesis of Geoffrey Waweru, a student in Physical Land Resources. The thesis titled Developing pedotransfer function to predict soil water retention curve of Kenyan soils under the direction of Prof. Dr. Ir. Wim M. Cornelis.

2012: Tutor of the MSc. thesis of Likeyesus Abebe, student in Physical Land Resources. The thesis titled 'Estimation of soil hydraulic properties from tension-disc infiltrometer data using an inverse method' under the direction of Prof. Dr. Ir. Wim M. Cornelis.

6.5. Reviewer's appointment by a scientific journal

African Journal of Agricultural Research

7. Software's ability

Microsoft Office (Word, Excel, PowerPoint, Access); SPSS; Matlab, R, SAS, SigmaPlot, Surfer.

8. Programming skills

Basic programming skills in Visual Basic for Applications in Excel, Matlab, R and SAS.

9. Languages

French (native), English (fluent) and Dutch (beginner).

10. Hobbies

Basketball, swimming, singing (choir member).

S