DEVELOPMENT AND EVALUATION OF SOIL WATER RETENTION PEDOTRANSFER FUNCTIONS FOR MEKONG DELTA SOILS IN VIETNAM

Nguyen Minh Phuong

GHENT UNIVERSITY

FACULTY OF BIOSCIENCE ENGINEERING

Promoter: **Prof. dr. ir. Wim M. Cornelis**

Department of Soil Management,

Faculty of Bioscience Engineering,

Ghent University, Belgium

Dean: **Prof. dr. ir. Mark Van Meirvenne**

Faculty of Bioscience Engineering,

Ghent University

Rector: **Prof. dr. Anne De Paepe**

Ghent University, Belgium

MSc. Nguyen Minh Phuong

DEVELOPMENT AND EVALUATION OF SOIL WATER RETENTION PEDOTRANSFER FUNCTIONS FOR MEKONG DELTA SOILS IN VIETNAM

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

Dutch translation:

ONTWIKKELING EN EVALUATIE VAN PEDOTRANSFERFUNCTIES VOOR HET VOORSPELLEN VAN BODEMWATERRETENTIE IN DE VIETNAMESE MEKONGDELTA

Please refer to this work as follows:

Nguyen, M. Phuong, 2016. Development and evaluation of soil water retention pedotransfer functions for Mekong Delta soils in Vietnam. PhD Thesis, Ghent University, Ghent, Belgium.

ISBN-number: 978-90-5989-940-7

*To the most wonderful people in my life***,**

My Family

ACKNOWLEDGEMENT

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

First of all, I would like to sincerely thank my promoter, Prof. dr. ir. Wim Cornelis for all his supports both in academic job and life. With the great scientific knowledge, he always gives me comprehensive guidance and constructive criticisms in working with scientific papers. Prof. Wim, thank you very much for always supporting me, particularly in some difficult periods in my PhD life.

I sincerely thanks the members of the examination committee: Prof. dr. ir. Pascal Boeckx (chairman), Prof. dr. Kristine Walraevens, Dr. ir. Henk Wosten, Prof. dr. ir. Piet Seuntjens, Dr. Jan Baetens, Prof. dr. Le Van Khoa for their thorough reading of my manuscript and their constructive comments and suggestions that undoubtedly helped to improve my thesis further.

I would like to give special thanks to colleagues and friends in the department of Soil Management, Ghent University who accompanied me for the past years. I am deeply grateful for their warm friendship, support and help. I would like to express my sincere thanks to Anita Lostrie, and Patrick Dosche for their help with administrative and other issues. Kindly thanks goes to the lab technicians Maarten Volckaert, Luc De Boosere, Mathieu Schatteman, Trees Vanbiervliet for their valuable helps in the lab works.

I gratefully acknowledge the funding received towards my PhD from the Ministry of Education and Training (MOET, project 322). I would like to express my deep appreciation to all of my colleagues at the Department of Soil Science, Can Tho University, Vietnam for their encouragement and support. Especially, I am very thankful to Prof. Dr. Le Van Khoa for all his help, advice, and encouragement.

I would like to thank all my Vietnamese friends in Ghent, especially to anh Toi, anh Thai, chi Hong Van, chi Bao Loan, chi Que, chi Ha, Anh Ngoc, Hai Giang, Thu Giang, Duc Tung, Loan Anh, Ut, Tho for their help and encouragement especially during the period of high pressure from work.

I am deeply thankful to my family: my parents, my sister’s family, my husband and my son who have always been with me to support and encourage me throughout my life with all their love, sacrifice and endless support. Without the encouragements from them, I don’t think I could have finished my PhD study.

TABLE OF CONTENTS

List of figures
List of tables
List of symbols
List of abbreviations
Summary
Samenvatting
Chapter 1. GENERAL INTRODUCTION 1
1.1. Problem statements 3
1.2. Potential predictors for soil hydraulic parameters estimation 6
1.3. Methods to develop PTFs 14
1.3.1. Statistical regression methods 14
1.3.1.1. Multiple linear regressions and polynomials of the n^{th} order 14
1.3.1.2. Extended nonlinear regression 15
1.3.2. Data mining or pattern-recognition techniques 16
1.3.2.1. Artificial Neural Networks (ANN) 16
1.3.2.2. Support vector machines (SVM) 19
1.3.2.3. k-Nearest Neighbor (kNN) 20
1.4. Type of soil water retention pedotransfer functions (SWRC-PTFs). 22
1.4.1. Class and continuous PTFs 22
1.4.1.1. Class PTFs 22
1.4.1.2. Continuous PTFs 23
1.4.2. Point-based PTFs, parameter-based PTFs, and pseudo-continuous PTFs 24
1.4.2.1. Point-based PTFs 26
1.4.2.2. Parameter-based PTFs 27
1.4.2.3. Pseudo-continuous based PTFs (PC-PTFs) 29
1.4.3. Semi-physically based PTFs and empirically-based PTFs 29

1.4.3.1. Semi-physically based PTFs 29
1.4.3.2. Empirically-based PTFs 31
1.4.4. Equation-based PTFs and pattern-recognition PTFs 31
1.5. Objectives of the study 31
1.6. Outline of the thesis 33
Chapter 2. DESCRIPTION OF THE STUDY AREA 35
2.1. Biophysical characteristics 37
2.1.1. Geographical location 37
2.1.2. Climate and hydrology 37
2.1.3. Geology 38
2.1.4. Vegetation and land use type 40
2.2. Socio-economic situation 41
2.3. Soils 43
2.3.1. Previous pedological studies about VMD soils 43
2.3.2. Soil samples used for pedotransfer modeling 47
2.3.2.1. PTFs' evaluation and calibration data set 47
2.3.2.2. PTFs' validation data set 49
2.3.2.3. Soil properties of the data sets used in this study 49
Chapter 3. PERFORMANCE EVALUATION OF PUBLISHED PTFS FOR PREDICTING SOIL WATER OF VIETNAMESE MEKONG DELTA SOILS 53
3.1. Introduction 55
3.2. Materials and methods 57
3.2.1. Evaluation data set 57
3.2.2. PTFs selected for evaluation 57
3.2.3. Evaluation methodology 61
3.2.3.1. PTFs' applicability 61
3.2.3.2. Evaluation criteria 61
3.3. Results and discussions 62

3.3.1. Exploratory data analysis 62

3.3.2. Applicability of investigated PTFs 63

3.3.3. Reliability of evaluated PTFs 64

3.3.4. Sources of prediction errors 71

3.3.4.1. Quality of PTFs’ calibration data sets 71

3.3.4.2. Development methods 72

3.3.4.3. Used predictors 73

3.4. Conclusions 73

Chapter 4. THE USE OF CATEGORICAL SOIL STRUCTURE INFORMATION IN DEVELOPING SOIL WATER RETENTION PTFS 75

4.1. Introduction 77

4.2. Materials and Methods 78

4.2.1. Soil data set 78

4.2.2. Predicting soil water retention 79

4.2.3. Evaluating the accuracy of prediction 79

4.3. Results and Discussions 80

4.3.1. Performance of point PTFs developed based on the whole dataset 80

4.3.2. Improved effects of grouping by categorical soil structural information on PTFs’ accuracy. 82

4.3.3. Effects of supplementary predictors in PTF’s predictability. 86

4.4. Conclusions 89

Chapter 5. COMPARISON OF DIFFERENT REGRESSION TECHNIQUES IN DEVELOPING PTFS 91

5.1. Introduction 93

5.2. Materials and Methods 95

5.2.1. Soil data sets 95

5.2.2. Types of soil water retention PTFs 96

5.2.3. Methods to derive PTFs 97

5.2.3.1. Multiple Linear Regression (MLR) 97

5.2.3.2. Artificial Neural Networks (ANN) 97
5.2.3.3. Support Vector Machines for Regression (SVR) 98
5.2.3.4. k-Nearest Neighbors (kNN) 98
5.2.4. Evaluation criteria 99
5.3. Results 100
5.3.1. Exploratory data analysis 100
5.3.2. Point PTFs performance 103
5.3.3. Pseudo-continuous PTFs (PC-PTFs) 107
5.4. Discussions 110
5.5. Conclusions 112
Chapter 6. COMBINED EFFECTS OF REGRESSION METHODS AND THE USE OF SOIL STRUCTURE ON PTFS' ACCURACY 115
6.1. Introduction 117
6.2. Materials and methods 119
6.2.1. Soil data set 119
6.2.2. Methods to build PTFs 122
6.2.2.1. Support Vector Machines 122
6.2.2.2. k-Nearest Neighbors (kNN) 122
6.2.3. Evaluation criteria 123
6.2.4. Statistical analysis 123
6.3. Results 124
6.3.1. Effect of soil structural grouping on SVM-PTFs 126
6.3.2. Effect of soil structural grouping on kNN-PTFs 127
6.4. Discussions 128
6.5. Conclusions 133
Chapter 7. FUNCTIONAL EVALUATION OF PTFS' PERFORMANCE USING AGRO-HYDROLOGICAL MODELS AQUACROP AND HYDRUS-1D 135
7.1. Introduction 137

7.2. Materials and methods: 139
7.2.1. Study area and soil data of investigated profiles 139
7.2.2. Simulation models 141
7.2.2.1. AquaCrop model 142
7.2.2.2. Hydrus-1D 145
7.2.3. Model inputs 147
7.2.3.1. Soil data 147
7.2.3.2. Climatic data and boundary conditions of the models 149
7.2.3.3. Crop data 150
7.2.4. Statistical evaluation of simulated time series 151
7.3. Results and Discussions 152
7.3.1. Comparison of SWRC 152
7.3.2. AquaCrop model 159
7.3.3. Hydrus 1D model 163
7.4. Conclusions 169
Chapter 8. GENERAL CONCLUSIONS AND RECOMMENDATIONS 171
8.1. The applicability of SWRC-PTFs for soils in VMD 173
8.2. Important strategies in PTFs development 175
8.2.1. Potential predictors of SWRC-PTFs 175
8.2.2. Regression methods 176
8.3. PTFs' utility for practical agro-hydrological modeling 179
8.4. Recommendations for future research 180
References 183
Appendices 209
Curriculum vitae 219

LIST OF FIGURES

Figure 1-1	Schematic overview of a three-layer neural network. Source: Schaap and Leij (1998).	17
Figure 1-2	Graphical explanation of the basic idea of Support Vector Machines (SVM) for a classification problem which display a non-linear decision boundary in original input space. After the training data is mapped into higher dimensional feature space, the linear separating hyper-plane is visualized and the decision surface with maximum-margin can be analytically found. Source: Shmilovici (2010).	19
Figure 1-3	Graphical representation of k-Nearest Neighbors (kNN) for finding the best match for a target soil. Source: Jagtap et al. (2004).	21
Figure 1-4	Four general categorization schemes of pedotransfer functions to estimate soil water retention characteristics (SWRC-PTFs).	22
Figure 1-5	The typical topologies of the point-based, parameter-based and pseudo-continuous pedotranfer functions when using an artificial neural network. Source: Haghverdi et al. (2012).	25
Figure 1-6	Flow-chart of the dissertation.	34
Figure 2-1	The possition of the Mekong Delta in Vietnam and it's corresponding admistrative map	37
Figure 2-2	Map of geology and physiography of the mainland of the Mekong Delta in Vietnam simplified after Mensvoort and Van den Berg (1986). Source: Le (2003).	40
Figure 2-3	The map of different soil groups in the Vietnamese Mekong Delta. Red stars represent the sampling sites of traing data set (i.e., data collected in the framework of this dissertation). Circles with number display the location of soil profiles of test data collected from the study of Le (2003).	46
Figure 2-4	Variation of soil texture classes in the training data set (grey circles) and test data set (black triangles).	51

Figure 3-1	Correspondence of measured vs. PTFs' predicted soil moisture content ($m^3 m^{-3}$) at -33 kPa (circles) and at -1500 kPa (crosses) of regression based PTFs developed for soils in temperate climates.	67
Figure 3-2	Correspondence of measured *vs.* PTFs' predicted soil moisture content ($m^3 m^{-3}$) at -33 kPa (circles) and at -1500 kPa (crosses) of regression based PTFs developed for soils in tropical climates.	68
Figure 3-3	Correspondence of measured *vs.* PTFs' predicted soil moisture content ($m^3 m^{-3}$) at -33 kPa (circles) and at -1500 kPa (crosses) of pattern-recognition based PTFs.	69
Figure 4-1	Scatter plot of measured water content ($m^3 m^{-3}$) versus PTFs' predicted water content ($m^3 m^{-3}$).	82
Figure 4-2	Variation of average RMSE values as a function of $\log_{10}h$ (h is the pressure head expressed in cm water) for point PTFs developed before and after grouping by using descriptive soil structural information.	85
Figure 4-3	Variation of average RMSE values as a function of $\log_{10}h$ (h is the pressure head expressed in cm water) for point-PTFs developed before (A) and after grouping by soil structure (B) using classical predictors set and extended predictors set.	88
Figure 5-1	Structural topologies of point PTFs, and pseudo-continuous PTFs (modified from Haghverdi et al., 2012).	96
Figure 5-2	Variation of root mean square error (RMSE) as a function of log(h) (h is matric head, expressed in cm water) in training phase and testing phase of point PTFs derived by multiple linear regression (MLR), artificial neural networks (ANN), support vector machines for regression (SVR), and k-nearest neighbors (kNN) methods.	105
Figure 5-3	Scatter plots of observed *vs.* predicted water content ($m^3 m^{-3}$) of point PTFs in calibration and validation phases. MLR-point-PTF, ANN-point-PTF, SVR-point-PTF, kNN-point-PTFs are the point PTFs derived by Multiple Linear Regression, Artificial Neural Networks, Support Vector Regression, and k-Nearest Neighbors techniques.	106

Figure 5-4 Scatter plots of observed vs. predicted water content (m^3 m^{-3}) of pseudo-continuous PTFs in calibration and validation phases. MLR-PC-PTF, ANN-PC-PTF, SVR-PC-PTF, kNN-PC-PTFs are the pseudo-continuous PTFs derived by Multiple Linear Regression, Artificial Neural Networks, Support Vector Regression, and k-Nearest Neighbors techniques. 109

Figure 6-1 Variation of soil texture of the data set (N=160) in the USDA textural triangle, where the circles, triangles and crosses represent soils belonging to massive, structured and structureless subgroups, respectively. 120

Figure 6-2 Variation of RMSE as a function of pF (pF = $\log_{10}(h)$, where h is the potential head expressed in cm of water) before and after grouping by soil structure for PTFs developed by Support Vector Machines (SVM) and k-Nearest Neighbor (kNN) approaches. 125

Figure 6-3 Variation of RMSE as a function of pF (pF = $\log_{10}(h)$, where h is the potential head expressed in cm of water) of specific subgroups (‘Mass’ is the massive soils group, ‘Stru’ is the structured soils group, and ‘StrL’ is the structureless soils group). SVM, kNN, SVM_Gr, kNN_Gr are the PTFs derived by Support Vector Machines and k-Nearest Neighbor approaches based on the whole data set and specific structural subsets, respectively. 125

Figure 6-4 Scatter plots of measured vs. SVM-PTFs’ predicted soil water content before and after grouping by categorical soil structure. The circles, triangles and crosses represent soils belonging to massive, structured and structureless subgroups, respectively. 127

Figure 6-5 Scatter plots of measured soil moisture content vs. kNN-PTFs’ predicted soil moisture content before and after grouping by categorical soil structure. The circles, triangles and crosses represent soils belonging to massive, structured and structureless subgroups, respectively. 128

Figure 6-6 Variation of soil water content (m3 m-3) at (a) air entry (pF 1), (b) field capacity (pF 2.5), and (c) permanent wilting point (pF 4.2) as a function of clay content, where the circles, triangles and crosses 129 represent soils belonging to massive, structured, and structureless subgroups, respectively.

Figure 6-7 Variation of soil water content (m^3 m^{-3}) at pF 1 as a function of bulk density, where the circles, triangles and crosses represent soils belonging to massive, structured and structureless subgroups, respectively 130

Figure 7-1 Daily precipitation and reference evapotranspiration ET_o calculated from FAO Penman Monteith equation for the two investigated rice growing seasons in 2011, i.e., (a) He Thu (HT) and (b) Dong Xuan (DX), with meteorological data from the meteorological station at Tien Giang province, Vietnam. 150

Figure 7-2 Soil water retention characteristics (SWRC) of three horizons of four studied profiles obtained from direct laboratory measurement (lab-derived SWRC) and indirectly predicted using locally-derived PTFs (VMD-PTFs), or globally provided PTFs (SWC-PTFs and Rosetta-PTFs). 153

Figure 7-3 Simulated rice yield in wet season and net irrigation requirement in dry season determined by AquaCrop model using SWRC data derived from different approaches (i.e., lab-derived SWRC, prediction via using VMD-PTFs and SWC-PTFs). 159

Figure 7-4 Daily variation of soil water storage (mm) in the root zone (Z_r = 0.5m) in the wet season of four studied profiles under different scenarios of SWRC inputs 161

Figure 7-5 Simulated daily moisture content distribution at three depths within the root-zone of four studied profiles under different simulation scenarios of using SWRC data determined by lab measurement (lab-derived SWRC) and via prediction with locally derived PTFs (VMD-PTFs) and globally offered PTFs (Rosetta-PTFs). 164

Figure 7-6 Simulated degree of water stress (DWS) in dry season of four studied profiles under different scenarios of SWRC inputs. For example: ST1, ST2, ST3 are the simulated DWS of ST profile when using SWRC parameters obtained from lab-measurement, VMD-PTFs and Rosetta-PTFs predictions, respectively. 167

LIST OF TABLES

Table 1-1	Succesful predictors used in soil hydraulic pedotransfer functions.	8
Table 1-2	Water retention at two matric potentials in samples from different textural classes of US National Pedon Characterization database (Soil Survey Staff, 1997). Source: Pachepsky and Rawls (2003).	24
Table 2-1	Major soil groups in the Vietnamese Mekong Delta.	45
Table 2-2	Descriptive statistics of soil properties in the training (N=160) and test (N=29) data sets.	50
Table 3-1	List of selected regression-based PTFs developed for soils from both tropical and temperate regions and used to test for the VMD soils.	59
Table 3-2	The range of soil attributes of the calibrated data sets used to derive the evaluated PTFs.	60
Table 3-3	Pearson correlation matrix between soil variables in the evaluation data set (N=160).	62
Table 3-4	Applicability index and statistical indices (ME, SDE, RMSE) yielded by the investigated PTFs.	66
Table 4-1	Pearson's correlation between potential predictors and measured soil volumetric water retention.	81
Table 4-2	Regression coefficients and coefficient of determination of point-PTFs.	81
Table 4-3	Predictors and coefficients of determination (R^2) of point-PTFs developed for different soil groups. Log(OC) is the logarithmic form of organic carbon content (%), BD is soil bulk density (Mg m^{-3}), Sa, Si, Cl are sand, silt, and clay content (%).	84
Table 4-4	Selected predictors and corresponding R^2 values for point-PTFs developed using the extended set of predictors.	87
Table 5-1	Pearson's correlation coefficients between soil properties in the training data set (N=160).	102
Table 5-2	Mean prediction error (ME) and coefficient of determination (R^2) of point PTFs derived based on various regression techniques, i.e., multiple linear regression (MLR), artificial neural networks (ANN),	104

support vector machines for regression (SVR), and k-nearest neighbors (kNN).

Table 5-3 Performance evaluation using different statistical indices as RMSE, ME and R^2 of point PTFs and PC-PTFs developed by multiple linear regression (MLR), artificial neural networks (ANN), support vector machines for regression (SVR) and k-nearest neighbors (kNN) methods. 107

Table 6-1 Descriptive statistics of soil properties in the whole training data set (N = 160 samples), and the specific subgroups of massive, structured and structureless soils (N = 91, 46, 23, respectively). 121

Table 6-2 Mean error (ME) and coefficient of determination (R^2) of point PTFs developed by Support Vector Machine (SVM), and k-Nearest Neighbors (kNN) techniques based on the whole data set (before grouping) and three structural subsets (after grouping). 124

Table 7-1 Characterization of soil physical, chemical and hydraulic properties of four studied profiles. 143

Table 7-2 Soil hydraulic inputs obtained from VMD-PTFs and SWC-PTFs of Saxton and Rawls (2006) for AquaCrop model simulation. 154

Table 7-3 Soil hydraulic parameters obtained via fitting the equation to point predicted SWRC using VMD-PTFs and via parameter-based PTFs of Schaap et al. (2001) in Rosetta software for Hydrus-1D model simulation. 155

Table 7-4 Outputs of soil water balance components obtained by simulations using inputs of (M1) lab-measurement, (M2) VMD-PTFs' prediction, and (M3) SWC-PTFs' prediction. 162

Table 7-5 The mean error (ME) of time series soil water content ($m^3\ m^{-3}$) simulated by using predicted data of VMD-PTFs and Rosetta-PTFs where simulation using lab-measured SWRC data was considered as benchmark for comparison. Numbers between the brackets represent the standard deviation of the error (SDE). 165

LIST OF SYMBOLS

Greek symbols

Symbol	Name	Unit
α	Scaling parameter of van Genuchten (1980) equation	kPa^{-1}
γ	Kernel parameter in SVM algorithm	
ε	Error	
ε	Precision parameter in SVM algorithm	
θ	Volumetric water content	$m^3\ m^{-3}$
θ_r	Residual volumetric water content	$m^3\ m^{-3}$
θ_s	Saturated volumetric water content	$m^3\ m^{-3}$
Ψ	Soil matric potential	kPa

Roman alphabet

a	Intercept	
B	Biomass	
d_i	Distance between estimation point and observation point at location x_i	m
h	Soil matric head	cm
h_b	Bubbling pressure	cm
i	Individual soil samples	
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	Number of observations/soil samples	
r	Coefficient of correlation	
R^2	Coefficient of determination	

w	Gravimetric water content	kg kg^{-1}
w_i	Weight assigned to soil *i*	
C	Regularization parameter in SVM algorithm	
$C(h)$	Specific moisture capacity	
z	Vertical coordinate	m
t	Time	day or hour
S	Sink or Source term in Richards' equation	
K_y	Crop yield response factor	
Z_r	Depth of root zone	m
S_p	Potential water uptake rate in the Feddes model	cm day^{-1}

LIST OF ABBREVIATIONS

Abbreviation	*Description*	*Unit*
AIC	Akaike Information Criterion	
ANN	Artificial Neural Networks	
ANOVA	Analysis of variance	
AWC	Available water capacity	$cm^3\ cm^{-3}$
BD	Bulk density	$Mg\ m^{-3}$
CART	Classification and Regression Trees	
CEC	Cation exchange capacity	$cmol_c\ kg^{-1}$
CV	Coefficient of variation	
DCB-Al	Dithionite-citrate-bicarbonate aluminium	
DCB-Fe	Dithionite-citrate-bicarbonate iron	
DWS	Degree of water stress	
E	Evaporation	
EC	Electrical conductivity	$mS\ cm^{-1}$
ERN	Extended non-linear regression	
ET	Evapotranspiration	
ET_0	Potential/reference evapotranspiration	
FAO	Food and Agriculture Organization	
FC	Field capacity	$cm^3\ cm^{-3}$
GA	Genetic Algorithms	
GAM	General Additive Model	
GLM	Generalized Linear Model	
GMDH	Group Method of Data Handling	
GP	Genetic Programming	
HAC	High activity clay	
HI	Harvest Index	
IGBP/T	Subset of tropical soils from the IGBP-DIS database	
IGBP-DIS	International geosphere biosphere programme data and Information System	
ISRIC	International Soil Reference and Information	

	Center	
K(h)	Unsaturated hydraulic conductivity	cm h^{-1}
kNN	k-Nearest Neighbors	
K_{sat}	Saturated hydraulic conductivity	cm h^{-1}
LAC	Low-activity clay	
LOO	Leave-one-out	
MARS	Multiple Adaptive Regression Splines	
MLR	Multiple Linear Regression	
OC	Soil oranic carbon content	%
OM	Soil organic matter content	%
PL	Plastic limit	kg kg^{-1}
POD	Pore size distribution	
PSD	Particle size distribution	
PTFs	Pedotransfer functions	
PWP	Permanent wilting point	cm^3 cm^{-3}
RETC	Retention Curve programme	
RMSE	Root mean squared error	
RT	Regression Trees	
SDE	Standard deviation of the errors	
SI	Stability index	
SQSP	Soil Quality Scoring Procedure	
SSE	Sum of squared errors	
SVM	Support Vector Machines	
SVR	Support Vector Machines for Regression	
SWRC	Soil water retention characteristics/curves	
T	Transpiration	
TAW	Total available water content	
UN	United Nations	
UNSODA	Unsaturated soil hydraulic database	
USDA	United States Department of Agriculture	
VESS	Visual Evaluation of Soil Structure	
VMD	Vietnamese Mekong Delta	
VSA	Visual soil assessment	
WP	Water productivity	

SUMMARY

In Vietnamese Mekong Delta (VMD), appropriate irrigation and drainage management, which requires information of soil water retention characteristics (SWRC), is key to sustainable paddy rice production. SWRC information, however, is usually missing due to the lack of facilities, cost, and personnel training involved in its direct measurement. Pedotransfer functions (PTFs) that provide estimations of SWRC from other basic soil properties are an alternative source of SWRC for practical soil water management or modeling purposes.

The main objective of this research was to develop and evaluate innovative SWRC-PTFs for tropical VMD soils in order to better understand the complex soil-water relationships in the studied region where recently developed alluvial soils have been mainly exploited for paddy rice production.

Since developing new PTFs is an arduous task which generally requires a large database of good quality, utilizing existing PTFs where possible is highly recommended. Such evaluation is very important because it would address the need of improving existing PTFs or developing new VMD-PTFs for regional application. In this dissertation, several well-known published PTFs which were developed for soils in both tropical and temperate climates were evaluated in terms of their applicability and reliability in predicting SWRC of VMD soils. The evaluation showed that the predictive performance of published PTFs for tropical VMD soils greatly varied among investigated PTFs and that it was more dependent on the coverage and the quality of the PTFs' training databases. The prediction errors by using 'tropical' PTFs were mainly attributed to difference in clay mineralogy, while unreliable predictions of 'temperate' PTFs resulted from the difference in the distribution of soil texture classes between training databases and testing data set. The findings suggested that specific SWRC-PTFs for tropical delta soils need to be developed for sustainable soil-water management in the region.

For developing new SWRC-PTFs, two important strategies in PTF research (i.e., identification of significant predictors of SWRC prediction and of appropriate regression techniques used in PTF development) were investigated in order to obtain the best-performing predictive models.

Determination of potential predictors for SWRC of recently developed alluvial soils in VMD was carried out using stepwise multiple linear regression (MLR). The results revealed that SWRC of tropical VMD soils could be satisfactorily estimated by classical PTF predictors (e.g., soil texture, BD, and OC). However, incorporating descriptive soil structural information (e.g., presence or absence of pedality) as grouping criterion prior to PTF development did improve the prediction accuracy of SWRC, especially in the wet moisture range. Plastic limit was also found to be a promising predictor for SWRC-PTFs of soils having a given degree of structural development.

To test the impact of regression methods, the predictive capabilities of point PTFs and pseudo-continuous (PC) PTFs developed by different regression techniques (i.e., Multiple Linear Regression (MLR), Artificial Neural Networks (ANN), Support Vector Machine for Regression (SVR), and k-Nearest Neighbors (kNN)) were compared and evaluated. The results showed that point PTFs derived by various regression techniques provided comparable accuracy, but the reliability of ANN, SVR and kNN models was much better than that of MLR. In case of PC-PTFs, ANN and kNN models outperformed SVR and MLR in both training and testing phases. Our findings confirm the superiority of data-mining approaches in modeling the complex system of soil and water even when a limited dataset is available (N = 160 observations) as the average prediction error (Root Mean Square Error, RMSE) for the test data varied from 0.049 to 0.053 m^3 m^{-3}.

The predictive performance of the PTFs was even more enhanced, though only slightly, by the combined effect of the two above-mentioned strategies, i.e., adding significant predictors as descriptive soil structural information, and implementing more flexible regression algorithms. Incorporating descriptive soil structure information improved the accuracy of PTFs derived by the SVR approach in the range of matric potentials of -6 to -33 kPa (average RMSE decreased up to 0.005 m^3 m^{-3} after grouping, depending on matric potentials). The improvement was primarily attributed to the outperformance of SVR-PTFs calibrated on data of the structureless soil group. No improvement was obtained with the kNN technique, at least not in our study in which the data set became limited in size after grouping.

Once the PTFs were developed, their utility was assessed through functional evaluation using two agro-hydrological models, i.e., the crop-water model AquaCrop and the hydrological model Hydrus-1D. Modeling water and solute transport has become an important tool in simulating agricultural productivity as well as environmental quality. In the functional evaluation, the effect of replacing direct laboratory measured SWRC (as reference) by PTF-predicted values (obtained from locally derived VMD-PTFs and widely-used existing PTFs) on the outcomes of the agro-hydrological models was investigated. The performance of agro-hydrological models is known to be sensitive to variation in soil hydraulic parameters. The results showed that for soils that are typically found in the VMD region (e.g. alluvial soil with clayey textured, acid sulfate soils with high OM content, and sandy soils), locally derived VMD-PTFs offer more accurate estimation of SWRC, and consequently also perform rather well (as compared with the simulation scenario using laboratory-measured SWRC data) in simulating different hydraulic responses of paddy fields in wet and dry conditions using both a conceptual crop-water model and a physically-based hydrological model. On the other hand, the results also raised awareness towards the arbitrarily application of widely-used PTFs in agro-hydrological models (e.g. PTFs of Saxton and Rawls, 2006, used in “Soil Water Characteristics” software and ANN-PTFs of Schaap et al., 2001 in “Rosetta” software) which may thus be a source of error in the model’s simulation.

To sum up, the applications of PTFs to soils which are different from or not represented by the data used to calibrate regression functions should be done with caution, particularly in hydrological model applications as their error in soil hydraulic properties estimation can jeopardize the modeling results of soil water balance components which in turn can affect management and planning guidance for natural resources management and agricultural production. Including pedological soil structure information into VMD-PTFs leads to better SWRC predictions. Well-tested PTFs, on the other hand, can be considered as a reasonable, reliable and cheap alternative to sampling and laboratory measurements of soil hydraulic data for modeling, especially in large-scale research projects. Further improvement of the predictive capacity of locally-derived PTFs can be made by increasing their coverage in terms of temporal and spatial variability of the soils.

SAMENVATTING

Een geoptimaliseerd irrigatie- en drainagebeheer van de rijstvelden in de Vietnamese Mekong Delta (VMD) is essentieel voor duurzame rijstproductie op de bevloeide akkers. Dit beheer vereist vochtretentiedata (SWRC) van de bodems, data die echter vaak ontbreekt wegens een gebrek aan infrastructuur, budget en opgeleid personeel. Pedotransferfuncties (PTF) die de vochtretentiecurve (VRC) schatten uit eenvoudig te bepalen bodemeigenschappen zijn een alternatieve bron voor VRC-data in bodem- en waterbeheer, o.a. als modelinput.

Het hoofddoel van dit onderzoek was het ontwikkelen en evalueren van innovatieve VRC-PTFs voor tropische bodems in de VMD, om zo een beter inzicht te krijgen de complexe bodem-vocht relaties in dit studiegebied waar jonge alluviale bodems gebruikt worden voor bevloeide rijstbouw.

Gezien het ontwikkelen van nieuwe PTFs een zware opgave is die een grote dataset van goede kwaliteit vereist, is het gebruik van bestaande PTFs waar mogelijk aangewezen. In deze studie werden verschillende gepubliceerde PTFs voor tropische bodems geëvalueerd naar toepasbaarheid en betrouwbaarheid voor het voorspellen van VRC in VMD om te bepalen of de ontwikkeling van een nieuwe PTF noodzakelijk is. De evaluatie toonde aan dat de prestaties van de bestaande PTFs onderling zeer sterk verschilden en dat deze verschillen grotendeels te wijten zijn aan de kwaliteit en grootte van de trainingdataset. Slechte schattingen waren vooral te wijten aan een verschil in kleimineralogie in het geval van tropische PTFs, en de verschillen in distributie van de bodemtextuurklassen in training- en testdataset voor gematigde PTFs. Hieruit werd geconcludeerd dat het aangewezen is om specifieke VRC-PTFs te ontwikkelen voor tropische alluviale bodems voor een duurzaam bodem- en waterbeheer in het studiegebied.

Voor het ontwikkelen van nieuwe VRC-PTFs werd onderzoek gedaan naar de identificatie van significante predictor-variabelen en naar de geschikte regressie-techniek. Met gebruik van stepwise multiple linear regression werden potentiële predictor-variabelen bepaald voor VRC in jonge alluviale bodems. De resultaten wezen uit dat VRC van de tropische VMD-bodems goed geschat werden met de klassieke PTF predictorvariabelen (bv. textuur, bulkdichtheid en organisch-koolstofgehalte), maar met inclusie van beschrijvende structuurinformatie (bv. de aan-

of afwezigheid van bodemstructuur). Het groeperen van de data in bepaalde structuurklassen bij de PTF-ontwikkeling zorgde voor een significante verbetering van de accuraatheid van de schattingen, in het bijzonder in de natte regio van de VRC (tussen zgn. veldcapaciteit en verzadiging). De plasticiteitsindex was ook een veelbelovende predictor-variabele voor bodems met een zekere bodemstructuur.

Om de impact van verschillende regressiemethoden te testen, werden de prestaties vergeleken van punt-PTFs en pseudo-continue (PC) PTFs, ontwikkeld met een gamma van regressietechnieken (Multiple Linear Regression (MLR), Artificial Neural Networks (ANN), Support Vector Machine for Regression (SVR), en k-Nearest Neighbors (kNN)). De resultaten gaven aan dat de punt-PTFs vergelijkbare accuraatheid behaalden, maar de betrouwbaarheid van ANN, SVR en kNN was veel beter dan MLR. Voor PC-PTFs werden veel betere resultaten behaald met ANN en kNN in de train- en testfase. Deze bevindingen bevestigen de superioriteit van data-mining technieken in het modelleren van de complexe relatie tussen bodem en water, zelfs indien slechts een beperkte dataset beschikbaar is (N = 160 observaties), met een gemiddelde root mean square error (RMSE) voor de test data die varieerde tussen 0.049 en 0.053 m^3 m^{-3}.

De prestatie van de PTFs kon zelfs een weinig verbeterd worden door het gecombineerde effect van de twee bovenvermelde strategieën: het toevoegen van significante predictorvariabelen, zoals beschrijvende structuurdata, en het implementeren van flexibelere regressie-algoritmen. Het incorporeren van beschrijvende bodemstructuur info verbeterde de accuraatheid van de SVR PTFs in de range van matrixpotentiaal tussen -6 kPa en -33 kPa (de gemiddelde RMSE verminderde tot 0.005 m^3 m^{-3}). De verbeterde prestaties zijn vooral te wijten aan accuratere schattingen voor structuurloze bodems. Met de kNN–techniek werd geen beter resultaat behaald, toch niet in deze studie, waar de dataset beperkt in grootte werd na groepering.

Eens de PTFs ontwikkeld waren, werd hun toepasbaarheid getest in een functionele evaluatie, gebruik makend van twee agro-hydrologische modellen: het gewas-watermodel AquaCrop en het hydrologische model Hydrus-1D. In deze evaluatie werd nagegaan wat het effect is van het vervangen van de gemeten VRC door de geschatte VRC op de output van de beide modellen. De sensitiviteit van agro-hydrologische modellen voor variaties in de hydraulische eigenschappen van bodems is gekend. De resultaten in deze studie toonden aan dat de simulaties voor typische

VMD bodems (alluviale bodems met kleitextuur en een hoog zwavelzuurgehalte en hoog organisch-materiaalgehalte, en zandige bodems) waarbij de VRC bepaald werd door de VMD-PTFs, vergelijkbare simulatieresultaten geven als de simulaties met gemeten VRC-data voor de hydraulische respons van rijstvelden, in natte en droge condities, zowel met Aquacrop als met Hydrus-1D. Dit in tegenstelling tot de simulaties met VRC op basis van PTFs voor gematigde bodems, waarbij de modellen minder goed presteerden. De resultaten geven aan dat het gebruik van wijdverspreide PTFs in agro-hydrologische modellen, zoals de PTF van Saxton en Rawls (2006) en ANN-PTFs van Schaap et al. (2001) in “Rosetta”-software, een bron van simulatiefouten kan zijn.

Samenvattend bevestigd deze studie dat men dient zich te hoeden voor het gebruik van PTFs voor bodems die verschillen van, of niet vertegenwoordigd zijn in de data die gebruikt werd om de PTF te ontwikkelen. Dit is vooral het geval in hydrologische modellen, gezien door fouten in de hydraulische bodemeigenschappen de nauwkeurigheid van het modelresultaat van de bodemvochtbalans beperkt is, hetgeen op zijn beurt het plannen en beheer van natuurlijke grondstoffen kan beïnvloeden. Het incorporeren van pedologische bodemstructuurdata in VMD-PTFs leidt tot verbeterde VRC-schattingen. Grondig geteste PTFs kunnen beschouwd worden als een valabel, betrouwbaar en goedkoop alternatief voor meetcampagnes en laboanalyses van hydraulische bodemeigenschappen voor modeltoepassingen, in het bijzonder voor studies op een grote schaal. Verdere verbeteringen van de schattingen van PTFs kunnen bereikt worden door het uitbreiden van de dataset om temporele en spatiale variabiliteit van bodems te capteren.

Chapter 1

GENERAL INTRODUCTION

1.1. Problem statements

Ensuring food security for an ever-increasing world population is one of the main challenges that society is facing today. According to the United Nations, there are 7.3 billion people on the planet to feed today and another 2 billion are expected to join by 2050 with much of the increase occurring in the developing countries of the tropics (UN, 2010). Taking the consequence of such projected growth population, by 2050, agricultural productions must increase by 60 percent globally, and by almost 100 percent in developing countries in order to meet the world's food demand alone (FAO, 2015). Such escalation in food production requires 19% increase of water consumption for both rain-fed and irrigated agriculture (UN-Water, 2013b).

The increase of food production has to come from higher yields through cultivating more crops and exploiting more arable land. However, realizing this is not an easy task. First of all, it is evidently perceived that increasing population concomitantly leads to the shrinking of per capita arable land as a result of conversion of arable land to nonagricultural uses, e.g. aquaculture, industry, social services and house constructions (Lal, 2000). Agricultural intensification, on the other hand, is strongly dependent on and thus limited by the capacity of the soil and the water to support crop production. As it is urgently noticed, in the tropics, where population growth is typically high (e.g. Asia, South America, Africa), agriculture intensification and mismanagement has depleted the soil fertility, polluted the surface and ground water, and jeopardized the function of natural resources to maintain food production in these areas for present and future generations (Lal, 2000; Linh et al., 2014; Linh et al., 2015a; Osman, 2013).

Serving as a staple food for about half of the world's population, rice is considered as one of the most important crops in terms of food supply and food security for more than 50 percent of the world's population. Globally, rice is grown on almost 155 million ha of the world's surface and produces approximately 661 million tons (in 2008) worldwide with more than 90 percent produced in Asia (Kögel-Knabner et al., 2010). Cultivated rice (*Oryza sativa*), as evolved from its semi-aquatic ancestor, is extremely sensitive to water shortage and is considered as a highly water consumable crop.

Analyzing global water scarcity, Rijsberman (2006) noted that water will be the major constraint for agriculture in the coming decades, as agricultural sector accounts for 70 percent of the global fresh water withdrawal. Until recently, in several regions like the Vietnamese Mekong delta (VMD) where densely distributed river canals were constructed for agricultural production, water usage for irrigation has been taken for granted. Such condition, however, cannot continue because the scarcity of water for agricultural production is becoming a major problem resulting from climate change and soil degradation. Indeed, climate change, which can no longer be neglected, has serious direct and indirect consequences for rice production, particularly in lowland deltas where rice has been intensively cultivated (Jalota et al., 2012). For example, sea level rise as a result of climate change causes inundation of land and thus a loss of agricultural land area. Another effect of inland transgression of the sea is salt intrusion into the soil, together with fresh water shortage for crop irrigation in dry season, making the land unsuitable for agriculture.

The threat of climate change, water shortage and soil degradation makes it important to create a synergetic management system which is able to support optimal crop growth and food supply without jeopardizing the quality of natural resources that are already under severe stress. Sustainable management of soil and water resources are expected to increase nearly 60 percent of world's food production (FAO, 2015) and are considered as the main strategies for mitigation of and adaptation to global climate change (UN-Water, 2013a; Vepraskas et al., 2009).

Soil and water are intimately related in nature; hence thorough understanding of their properties, behaviors and interactions is essential for their rational management. Sustainable management of soil and water resources requires a collaborative effort of various soil and water related disciplines and expertises (Botula, 2013), which help to improve our understanding of complex soil and water interaction affecting rational crop production and agricultural practices (Bouma et al., 1999).

In recent years, there exists a growing interest to bridge traditional pedology with hydrology and soil physics in order to address diverse soil and water issues at various spatial and temporal scales (Lin et al., 2006). The emerging interdisciplinary research domain of hydropedology, bridging pedology and hydrology, is gaining substantial attention. Such interaction is desirable because the wealth of pedological information can advance the understanding and prediction of water distribution in soils and landscapes, whereas advances of hydrology can enrich interpretation of spatial distribution of soil properties (Pachepsky et al., 2006). The interdisciplinary domain of hydropedology is therefore expected to considerably contribute to our understanding of a wide range of agricultural, ecological, natural resource and engineering problems of societal importance.

Pedotransfer functions (PTFs) typically translate the *pedological* data we have, e.g. available from soil surveys, into the *hydraulic* data we need, i.e. soil hydraulic properties that govern water retention and water flow in soils (Wösten et al., 2001). Accurate estimates of soil hydraulic properties are required for running a wide range of models to simulate transport of water and solutes in the soil for a variety of applications ranging from field-scale water flow studies to global climate change (Twarakavi et al., 2009). In particular, with the increasing popularity of Geographical Information Systems (GIS) coupled with agro-hydrological models, functions that accurately estimate soil hydraulic properties are urgently needed. PTFs' knowledge and data are used for multiple applications of soil research in environmental science and engineering such as soil water flow modelling, predictive soil mapping, or filling the gaps in soil data sets (Pachepsky et al., 2015)

Specifically, in the Mekong Delta in Vietnam, irrigation and drainage techniques are crucial to maintain the soil moisture within a desirable range. Due to its low-lying position and typical monsoon climate, the Mekong Delta suffers from both water surpluses in the rainy season and water shortage in the dry season. Whereas in the rainy season as much as half of the plain is immerged by rain and floods, there is not enough fresh water for agricultural production in several areas in the dry season (Le, 2003). Moreover, the effects of global climate change (e.g. higher average temperatures and changes in rainfall distribution, soil moisture, river and groundwater flows (Mainuddin et al., 2013)), and the exploitation of rice-shrimp or intensive shrimp farms in the coastal area have led to the intrusion of saline water into the cultivated fields. Together with the increased use of agrochemicals, this inevitably impairs soil quality, pollutes the surface and ground water, degrades the environment and declines the functions of natural resources in supporting ecosystems and human life, unless appropriate soil-water management measures are taken. Such measures, and in particular irrigation and drainage practices, result in increased food production, raise income and boost economic development, while safeguard, restore or improve the physical, chemical and biological quality of the soil.

Depending on the complexity of the system, simple FAO-type of modelling tools to more sophisticated numerical models based on the Richards equation need to be implemented for developing sustainable irrigation and drainage systems (Bastiaanssen et al., 2007). The required estimations of soil hydraulic properties are needed as direct input in case of the simpler (bucket) models (e.g. AquaCrop, Steduto et al. (2009)), or as initial estimates when calibrating the more sophisticated numerical models (e.g. SWAP, van Dam et al. (2008); and HYDRUS, Šimůnek et al. (2008)).

As a tool for packaging and disseminating the soil information essential for environment prediction and risk assessment, PTFs are increasingly in demand in research related to global change. In fact, PTFs are indispensable in data-poor environments and large scale projects. However, only limited number of PTF studies has been conducted with international databases of soils from the humid tropics (Minasny and Hartemink, 2011; Tomasella and Hodnett, 2004). As it is indicated by Minasny and Hartemink (2011), very few efforts are devoted in prediction of hydraulic 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 out-dated. Consequently, there exist a certain number of knowledge gaps related to pedotransfer modeling of hydrophysical properties of soils of the humid tropic, particularly paddy soils in tropical delta, and its applications for sustainable land management (Botula, 2013).

1.2. Potential predictors for soil hydraulic parameters estimation

Factors affecting water retention and transport of water and chemicals in soils are manifold ranging from basic soil physical (soil texture, soil structure, bulk density), chemical (OC, cation exchange capacity - CEC), mineralogical (low activity clay and high activity clay) and mechanical properties (e.g., soil penetration resistance, shear strength, plasticity index) to bio-geo-morphological properties such as plant characteristics, soil management and topographical attributes of landscape

(Wösten et al., 2001). The potential predictors used in soil hydraulic PTFs can be obtained from different sources of information such as laboratory analytical data, field survey and soil morphological description, and the soil electromagnetic spectrum from laboratory samples as well as from field measurement with proximal and remote sensors (Minasny and Hartemink, 2011). **Table 1-1** lists the soil properties used most often in soil hydraulic PTFs because of their availability or because they proved to be promising. The role of different soil properties used as predictors in hydraulic PTF development can, according to the review of Wösten et al. (2001) and Vereecken et al. (2010), be briefly summarized as follows:

Table 1-1. Succesful predictors used in soil hydraulic pedotransfer functions.

Sources of data	Successful predictors	References of used PTFs
Laboratory analysis	*Particle/Aggregate size properties*	
	Sand, silt, clay content	Aina and Periaswamy (1985); MacLean and Yager (1972); Rajkai and Várallyay (1992); Saxton et al. (1986); Williams et al. (1992); Wösten et al. (1999)
	Detailed soil particle size distribution	Arya and Paris (1981); Børgesen et al. (2008); Haverkamp and Parlange (1986); Hwang and Choi (2006); Vereecken et al. (1989)
	Median or geometric mean of particle sizes	Giménez et al. (2001); Minasny et al. (1999); Scheinost et al. (1997)
	Aggregate size distribution	Guber et al. (2004)
	Hydraulic characteristics	
	Water content at -33 kPa and/or -1500 kPa	Børgesen et al. (2008); Botula et al. (2013); Saxton and Rawls (2006); Schaap et al. (1998); Schaap et al. (2001)
	Reference moisture retention curve	Rawls et al. (1998)
	Other physical, chemical, and mineralogical properties	
	Bulk density	Rawls and Brakensiek (1982); Rawls et al. (1998); Saxton and Rawls (2006); Schaap et al. (2001)
	Organic matter content	Rawls and Brakensiek (1982); Rawls et al. (2004); Rawls et al. (2003); Wösten et al. (1999)
	Cation exchange capacity (CEC)	Bell and van Keulen (1995); Botula et al. (2013); Hodnett and Tomasella (2002)
	Calcium carbonate ($CaCO_3$)	Khlosi et al. (2013); Rajkai and Várallyay (1992)
	Iron and/or aluminum oxides	Botula (2013); van den Berg et al. (1997)
	Clay type	Ali and Biswas (1986); Gaiser et al. (2000)
	Mechanical properties	
	Penetration resistance	Bayat et al. (2013); Giménez et al. (2001); Pachepsky et al. (1998)
	Atterberg limits (e.g. plastic limit)	Khlosi et al. (2013); Rawls and Pachepsky (2002)
Field and soil	Taxonomic information	Williams et al. (1983)

morphological description	(genetic horizons, parent material)	Jana and Mohanty (2011); Sharma et al. (2006)
		Leij et al. (2004); Pachepsky et al. (2001); Romano and Chirico (2004); Romano and Palladino (2002)
	Visible macro porosity	Anderson and Bouma (1973); McKenzie and Jacquier (1997)
	Soil structure description (i.e., grade, size, shape)	Lilly and Lin (2004); Pachepsky and Rawls (2003); Rawls and Pachepsky (2002); Williams et al. (1983)
	Soil consistence	Pachepsky et al. (1998)
	Pedality	Williams et al. (1983)
	Vegetation	Jana and Mohanty (2011); Sharma et al. (2006)
	Terrain attributes	Leij et al. (2004); Pachepsky et al. (2001); Romano and Chirico (2004); Romano and Palladino (2002)
Soil electromagnetic spectrum	Mid-infrared spectroscopy	Janik et al. (2007); Minasny et al. (2008b); Tranter et al. (2008)
	Visible – near infrared spectroscopy	Babaeian et al. (2015); Kodaira and Shibusawa (2013)

- *Soil texture or particle size distribution* is widely recognized as basis predictor variable for estimating soil hydraulic parameters and is used in almost any soil hydraulic PTF. Several researches, e.g., the one of Arya and Paris (1981); Gupta and Larson (1979); Rawls et al. (1998); Williams et al. (1983), have successfully verified the significant relationships between soil water characteristics and soil texture parameters. Different national and international classification systems, however, use quite different particle size classes, hence textural classes used in PTFs vary considerably (Wösten et al., 2001). Nonetheless, the use of sand, silt and clay content is common in many PTFs researches (Gupta and Larson, 1979; MacLean and Yager, 1972; Williams et al., 1983). On the other hand, many authors (e.g., Arya and Paris (1981); Haverkamp and Parlange (1986); Hwang and Choi (2006)) have proved that using more detailed particle size distribution data could improve the PTFs' accuracy. Moreover, it has been found that for soils with a wide texture range, median diameter of different particle size fractions is more useful to characterize particle size distribution. Hence, together with geometric standard deviations, median diameters were used by Minasny et al. (1999); Scheinost et al. (1997) to predict soil water retention, and by Mishra et al. (1989) to estimate saturated hydraulic conductivity.

- *Porosity or bulk density* is also important soil variables used in PTFs development. As it is widely proved, soil hydraulic properties are controlled not only by soil particle properties (soil texture) but also by their spatial arrangement and organization to form a porous medium in which soil water is being held and transported (i.e. soil structure) (Or and Wraith, 2002). The structure of single-grains soils, as well as aggregated soils can be considered quantitatively in terms of total porosity and pore size distribution. Tranter et al. (2007) proposed a conceptual model that considers BD as the result of particles packing and soil structure, hence the value of BD gives an indication of total soil porosity and is considered as proxy for soil structure quantification in PTF development. Indeed, Pachepsky and Rawls (2003) indicated that BD is a measurable continuous variable that is indirectly related to soil structure. There are many studies that used BD, beside soil texture and/or OM, as a predictor in predictive equations, for example, the PTFs of Aina and Periaswamy (1985); Rajkai and Várallyay (1992); Rawls et al. (1982); Rawls et al. (1998); Wösten et al. (1999).

- *Limited water retention data* at one or two points on the SWRC, usually taken at FC (-33 kPa) and PWP (-1500 kPa), have been used in several PTFs (Botula et al., 2013; Rawls et al., 1982; Saxton and Rawls, 2006) and improved considerably the accuracy of soil water retention PTFs. Such improvement is probably explained by the fact that soil water retention data provide more information about pore soil structure than texture and bulk density.

- *Mineralogical properties* such as the proportion of montmorillonite, kaolinite or illite clay, or the proportion of soils contained low-activity clay (LAC) (defined by CEC values) or high-activity clay (HAC), were shown to provide important information which could improve PTFs' accuracy (Ali and Biswas, 1986). Clay mineralogy of the soil is responsible for the structural development and porous behavior of the soil, beside the retention of water by absorption (Botula et al., 2013). As it has been proved in literature, low-activity clays such as kaolinite and halloysite which generally occur in highly weathered soils (e.g. Acrisols and Ferrasols) have low CEC values and low water retention capacity (particularly at high matric potentials), whereas high-activity clays such as montmorilonite present in swelling-shrinking soils (e.g. Vertisols) have high CEC value and high soil water retention capacity and poor internal drainage (Gaiser et al., 2000). Indeed, Hodnett and Tomasella (2002) found that CEC can be a promising predictor of parameter-based PTFs of soils in the tropics. Gaiser et al. (2000), on the other hand, used soil mineralogical information (i.e. LAC and non LAC) as grouping criterion to develop different PTFs for soils showing similar soil water retention behavior.

- *Organic matter/carbon content* was successfully used as input variable for SWRC estimation in numerous studies (e.g. Rawls and Brakensiek (1982); Rawls et al. (1982); Rawls et al. (2003); Wösten et al. (1999)). Generally, increased OM produces a soil with increased water holding capacity and conductivity, largely as a result of its influence on soil aggregation and associated pore space distribution (Saxton and Rawls, 2006). Water content at high tension, e.g. -1500 kPa, is determined principally by soil texture, thus, there is minimal influence by aggregation and OM. The effects of OM changes for higher soil-water contents vary with soil texture, particularly clay. OM effects are similar to those of clay, thus for soils with high clay content, the effects of increased OM on soil hydraulic properties are usually masked. Moreover, several studies have also reported a strong interrelation between OM and BD, and suggested that the effect of OM on soil water retention can be accounted for indirectly by using BD only (Manrique and Jones, 1991; Rawls et al., 2003; Zacharias and Wessolek, 2007) .

- *Other chemical properties such as extractable Fe and Al, and* $CaCO_3$ are also potential predictors for soil water retention estimation (Wösten et al., 2001) as they act as binding agents to form soil structure. The presence of Fe- and Al-oxides, and hydroxides in soils were also considered as indicators of the soil weathering state (Botula et al., 2012). Baert (1995) explained that in the ultimate stage of weathering, the break-down of both unstable primary and 2:1 minerals becomes complete and soils have reached 'the mineralogical weathering residue system' predominated mostly by Al- and Fe-(oxide) hydroxides and kaolinite (mineralogical maturity). These chemical substrates act as binding agents to form stable micro-aggregate structure of highly weathered soils (Botula et al., 2012; van den Berg et al., 1997), hence these soil variables are helpful in explaining the variation of soil water retention characteristics in the humid tropical regions. Moreover, $CaCO_3$ content can be as another important variable for estimating soil water retention, particularly in arid and semi-arid areas. Rajkai and Várallyay (1992) found the $CaCO_3$ content to be the second most important PTF input to predict water content at -1500 kPa for soils in

Hungary. This finding was confirmed by Khlosi et al. (2013) for Syrian soils. They found that the procedure of eliminating $CaCO_3$ in laboratory determination of soil texture lead to a significant mismatch between soil textures determined in the laboratory and that from tactile field description. Non-destruction of $CaCO_3$ in laboratory determination of soil texture increased the correlation between texture and hydraulic properties, and as a consequence, improved texture-based PTFs' predictive capacity.

- *Topographical/terrain attributes* were suggested to be used as potential predictors, along with basic soil properties, in PTFs (Rohdenburg et al., 1986). Incorporation of topographic attributes, e.g. elevation, slope, aspect, and potential solar radiation, helped improving the accuracy of original PTFs derived with basic soil properties in characterizing the spatial variability of soil hydraulic properties in a hill slope in Basilicata, Italy (Leij et al., 2004). Pachepsky et al. (2001) showed a good relationship of water retention with slope and curvature of soil surface. Romano and Palladino (2002) showed that the use of terrain attributes could correct systematic biases in water retention predictions from PTFs which originally used only basic soil predictors. Their results confirmed the role of terrain variables in representing the spatial patterns of soil hydraulic characteristics. Similarly, through developing hydraulic PTFs for soils in the Southern Great Plains of the USA, Sharma et al. (2006) concluded that in addition to pedological attributes, the use of topographic and vegetation attributes available from digital elevation models (DEMs) and normalized difference vegetation index (NDVI) did improve PTFs' performance.

- *Mechanical properties and shrink-swell parameters* as characterized by the coefficient of linear extensibility was used to predict both water retention (Pachepsky et al., 1998) and K_{sat} (Mckenzie et al., 1991). The plastic limit, which describes the soil-water content at which a soil begins to crumble when rolled into small threads, is also a promising predictor of soils water retention PTFs (Rawls and Pachepsky, 2002). Recently, Khlosi et al. (2013) investigated the role of soil plastic limit in estimating soil water retention of Syrian soils and confirmed that it was very promising for PTF developments.

- *Soil structure and morphology descriptors* are generally beneficial to PTF development. An assessment of soil structure can be done through (1) qualitative description of soil morphology in the field (e.g., presence of pedality, shape and size of structural units, grade of structural development), and/or (2) quantitative measurement of soil physical properties (e.g., soil penetration resistance, bulk density, total porosity) to characterize the soil matrix. Williams et al. (1983) included the field attribute pedality in soil water retention PTFs. They found that morphological structural properties (pedality, grade of structural development) were most consistently associated with soils having similar soil water characteristics. Danalatos et al. (1994) suggested to use grade of structural development as grouping criterion in PTFs development. A detailed measurement and count of lengths and widths of voids allowed Anderson and Bouma (1973) to estimate soil hydraulic conductivity of an argillic horizon of silt loam soils. Lin et al. (1999a, b) presented an elaborated system of morphometric indices and showed that these indices appeared to be the best predictors of micro- and macro-flows. Since the climate, crop, and human activities together have strong influence on soil structure developments, Lilly et al. (2008) believed that the use of soil structure information would help improve the effectiveness or the portability of hydraulic PTFs to other bio-climatic conditions. Working with soils from both temperate and tropical climates, Pulido Moncada et al. (2014) found that including morphological soil structural description would improve the prediction of soil saturated hydraulic conductivity. Furthermore, in a recent review on using PTFs for estimating hydraulic parameters, Vereecken et al. (2010) concluded that further improvement of PTFs could be achieved by adding new information such as soil structure and questioned how to best include soil structural information into PTFs.

- *Soil management* in the form of no-till resulted in an observed soil water retention to be larger than in conventional tilled soil in the range of capillary pressure of -30 to -400 kPa as presented by Azooz et al. (1996). Comparison of pre- and post-tillage shapes of SWRC appears to be indicative for changes in hydraulic characteristics (Klute, 1982). If tillage operations produce an increase in BD of a soil with a unimodal PSD, then K_{sat} will decrease 2-5 times. On the other hand, if tillage operations create a bimodal PSD, it is expected that both K_{sat} and water retention close to saturation will increase. Reynolds et al. (2007) reported that agricultural soil managements affect several near-surface soil physical (e.g. BD, macro-porosity, etc.) and hydraulic properties (relative water capacity and K_{sat}).

1.3. Methods to develop PTFs

There are various methods that could be employed to derive SWRC-PTFs. Although the soils in the tropics are different from soils in the temperate regions in several aspects, the general principles of soil-water relationship are similar (Minasny and Hartemink, 2011). Hence, various methods which were successfully applied to develop PTFs in the temperate regions are undeniably capable for PTFs in the tropics, although calibration and selection of important predictors remains necessary.

Generally, two main categories of methods for PTF's development can be distinguished: statistical regression approach including linear and non-linear models, and data mining and pattern-recognition techniques (Botula et al., 2014; Pachepsky and Schaap, 2004).

1.3.1. Statistical regression methods

Statistical regression is concerned with the analysis and construction of a dependence structure between dependent or response variables (e.g., soil-water content at specific matric potentials or parameters of analytical equations describing the whole SWRC) and independent or predictor variables (e.g., textural information, bulk density and organic matter content). A major part of the available and well-established PTFs for the prediction of soil hydraulic properties from continuous soil properties is based on statistical regression (Vereecken and Herbst, 2004).

1.3.1.1. Multiple linear regressions and polynomials of the n^{th} order

Concerning to statistical regression based PTFs; they either constitute multiple linear regression or polynomials of n^{th} order equations. The multiple linear regression technique is the mainstay of statistics for regression over the past 30 years and still remains one of the important tools used for prediction problems (Hastie et al., 2009; Vereecken and Herbst, 2004). The multiple linear regression equation for the prediction of the response variable y from a number of n predictor variables x_i can be written as:

$$y = a + \sum_{i=1}^{n} b_i x_i + \varepsilon \tag{1-1}$$

with constant intercept a, regression coefficients b_i, the error ε.

The predictability of the linear model depends mostly on the properties of the training database. For some scenarios, a linear model is the best one in describing the relationship between input and response variables, and hence the estimation of MLR is almost optimal. However, in the cases that the relationship is actually not linear and disjoint, linear models are unlikely to be optimal and other non-linear regression equations might be more adequate to describe the structure of such relationship.

A non-linear regression equation based on a second-order polynomial has the following form:

$$y = a + \sum_{i=1}^{n}(b_i x_i + c_i x_i^2) + \varepsilon \qquad (1\text{-}2)$$

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).

For examples, Gupta and Larson (1979) using MLR technique to derive point PTFs in the following form:

$$\theta_p = a(\%\ Sa) + b(\%\ Si) + c(\%\ Cl) + d(\%\ OM) + e(BD,\ Mg\ m^{-3}) \qquad (1\text{-}3)$$

where θ_p is volumetric soil water content ($m^3\ m^{-3}$) at different matric potentials *p*; Sa, Si, Cl are the percentage of sand, silt and clay fractions, OM is organic matter content, and BD is bulk density; *a, b, c, d, e* are regression coefficients of correspondent predictors (i.e., Sa, Si, Cl, OM, BD). PTFs at different matric potentials have different values of regression coefficients.

Most of the aforementioned point-based and parameter-based PTFs developed for soils in both temperate and tropical regions are MLR PTFs (e.g., Minasny and Hartemink (2011); Rawls and Brakensiek (1982); Tomasella and Hodnett (1998); Vereecken et al. (1989)), only some PTFs (e.g., Tomasella et al. (2000) and Minasny and Hartemink (2011) for the prediction of PWP) are polynomials of the n^{th} order.

1.3.1.2. Extended nonlinear regression

For developing parameter-based PTFs, Scheinost et al. (1997) found difficulties in estimating the scaling (*α*) and shape (*n*) parameters of the van Genuchten (1980) equation using the linear regression approach. They proposed an approach which first (1) sets up the expected relationship between the parameters of the hydraulic model and the soil properties; and then (2) inserts the relationship into the hydraulic model and estimates the parameters of the relationship simultaneously by fitting now the extended model to all data using nonlinear regression. This approach is referred to as extended nonlinear regression (ENR) by Minasny et al. (1999). Using soils from Australia, they compared MLR and ENR approaches in developing 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 was used by Hodnett and Tomasella (2002) to develop PTFs for the parameters of the van Genuchten (1980) equation.

Briefly, statistical regression techniques offer simple, reasonable and well-interpretable models. However, they also expose several drawbacks such as heavily biased estimations in case of small sample size. The right form of the regression equations which is usually unknown has to be determined *a priori*, rigorous assumptions about probability distribution of errors are not easy to fulfill across the data space, and the regression equations need to be redeveloped and republished in case new data become available (Botula et al., 2013; Nemes et al., 2006a; Patil et al., 2013).

1.3.2. Data mining or pattern-recognition techniques

Several new regression methods belonging to data mining or pattern recognition techniques have been introduced as a promising tool for PTF development, such as: ANN, Classification and Regression Tree (CRT), GMDH, GP, kNN, SVM, among others. The general description of these methods can be inferred in the review of Botula et al. (2014). In this section, only the most widely used data mining techniques such as ANN, SVM, and kNN techniques are presented. Due to their high flexibility and promising empirical performance, these data mining techniques have gained popularity in PTF research nowadays.

1.3.2.1. Artificial Neural Networks (ANN)

Besides the standard statistical regression methods, ANN is one of the earliest data mining techniques which have been applied to derive PTFs (e.g., the works of Pachepsky et al. (1996b); Schaap et al. (1998)). ANNs are intelligent machines, working very similar to an animal brain. An ANN consists of many interconnected simple computational elements called nodes or neurons (**Figure 1-1**). The artificial neural network can process very non-linear and complex data; hence the behavior of complex systems as soil water can be successfully simulated by varying the strength of the influence of network components to each other, and by varying the structure of the interconnections among components. After establishing network structure and finding coefficients to express the strength of influence of the network components to each other, an ANN becomes a complex formula of special type, relating input values with output values (Alexander and Morton, 1990), which can be then used like a regression function.

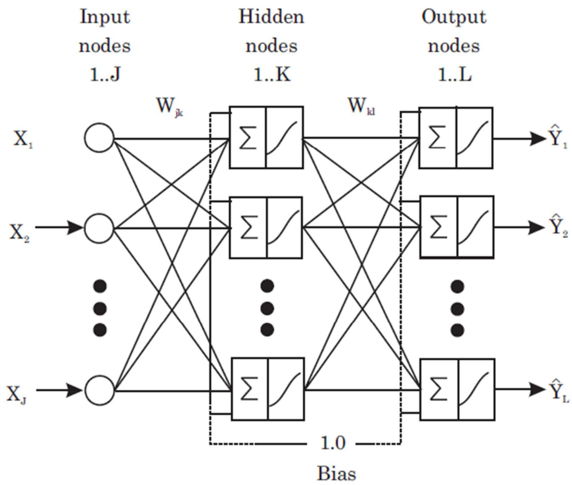

Figure 1-1. Schematic overview of a three-layer neural network.
Source: Schaap and Leij (1998).

The ANN technique was successfully used by Koekkoek and Booltink (1999) to predict water retention at various matric potentials based on Dutch and Scottish soil data sets. At the same time, using an extensive database of soils in the USA, Schaap et al. (1999) developed ANN PTFs to determine the parameters of the van-Genuchten (1980) equation. Because of the good performance of ANN PTFs, Schaap et al. (2001) introduced the ROSETTA software, a computer program that implements four hierarchical ANN PTFs for the estimation of the van Genuchten (1980) parameters and is embedded as pedotransfer functions in the physically-based water and solute transport model HYDRUS, as ROSETTA Light. The stand-alone ROSETTA software combines neural network analyses with the bootstrap method (i.e. drawing of random subsamples from an original sample through a sampling with replacement technique) (Efron and Tibshirani, 1993), thus allowing the program to provide uncertainty estimates of the predicted hydraulic parameters.

Minasny and McBratney (2002) proposed a new objective function for parameter-based ANN PTFs. The authors argued that this new method, called the neuron-m method, provides 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. Inheritably, Sharma et al. (2006) developed ANN models in a combination with the bootstrapping technique to predict moisture content (at eight different matric potentials) and the van Genuchten (1980) parameters.

The major advantage of ANN PTFs, as compared to MLR PTFs, is that they do not require a priori concept or assumption about the relations between input and output data. Such relationships in general are difficult to define, because these models are not known (Skalová et al., 2011). The training of an ANN is basically an iterative process, and during an iterative calibration procedure, the optimal relations between input and output data are found and implemented automatically (Schaap and Leij, 1998).

Despite the promising empirical performance, some problems from ANN characteristics may sometime arise. Twarakavi et al. (2009) summarized several possible weaknesses of the ANN approach when it is used to develop PTFs, such as: (1) ANN has a number of coefficients (weights) that do not permit easy physical interpretation (Schaap et al., 2001); (2) the ANN's structure has to be selected a priori and therefore may not be optimal since there are many types of neurons and many types of possible connections (Wösten et al., 2001); (3) a higher number of neurons and connections than required can result in over-fitting and over parameterization (Hastie et al., 2009); and (4) due to complexity of ANN's structure and the large number of weights that are being 'trained' as the network 'learns', there is no assurance that the learning algorithms will find optimum weights that minimize prediction errors.

The ANN PTFs have rarely been used to develop SWRC-PTFs for tropical soils (Botula et al., 2014). Some studies used the ROSETTA program to derive the parameters of the van Genuchten (1980) model for shrink-swell and highly weathered soils, and compared the results with locally-derived or published PTFs based essentially on MLR techniques (Botula et al., 2012; Patil and Rajput, 2009).

1.3.2.2. Support vector machines (SVM)

Support Vector Machines (SVM) are statistical learning machines applicable to both classification and regression problems. Originally, the concept of the SVM classifier was introduced by Vapnik (1995) to determine a maximum-margin hyper-plane that lies in a transformed input space and split the example classes while maximizing the distance to the nearest cleanly split examples. The parameters of the solution hyper-plane are derived from a quadratic programming optimization problem. **Figure 1-2** provides an example of the SVM's idea for a classification problem.

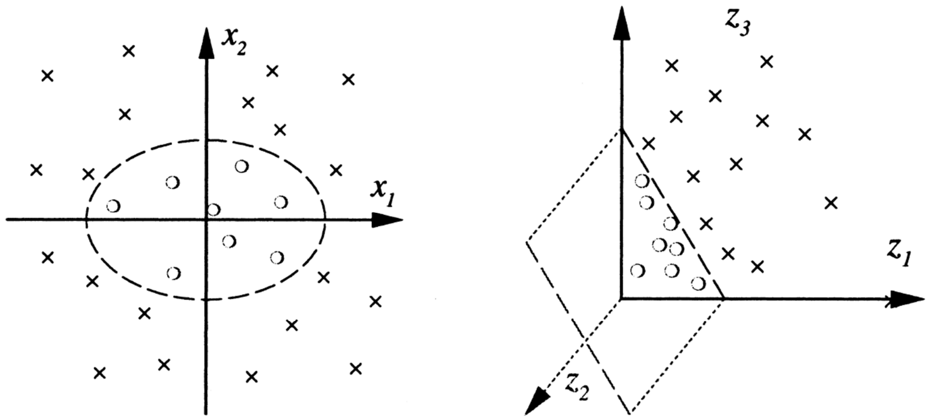

Figure 1-2. Graphical explanation of the basic idea of Support Vector Machines (SVM) for a classification problem which display a non-linear decision boundary in original input space. After the training data is mapped into higher dimensional feature space, the linear separating hyper-plane is visualized and the decision surface with maximum-margin can be analytically found. Source: Shmilovici (2010).

In the same way as with classification approach, SVM for regression (SVR) still contains all the main features that characterize a maximum-margin algorithm (Smola and Schölkopf, 2004). The basic idea of SVM for regression is to project the input data by means of kernel functions into a higher dimensional kernel induced feature space, where a linear regression can be performed for an originally nonlinear relation, the results of which are then mapped back to the input space. The linear regression is maintained by quadratic optimization, which ensures a global optimum of the SVM model. Moreover, instead of minimizing the observed training error as in statistical regression techniques, SVM formulation attempts to minimize the generalization error bound to achieve generalized performance through implementing the structural risk minimization principle in addition to the traditional empirical risk minimization. These features make SVM more capable than ANN to overcome local minimum and over-fitting problems, hence it is becoming more popular in many fields traditionally dominated by ANN (Lamorski et al., 2008; Vapnik, 1995).

Due to a promising ability to generalize the prediction, the SVM approach has attracted greater interest recently in agricultural and biological engineering (Skalová et al., 2011). Lamorski et al. (2008) used SVM to predict water retention of Poland soils at eleven matric potentials using sand, silt, clay, and BD as input predictors. Twarakavi et al. (2009) used the same algorithm to derive hierarchical SVM PTFs for predicting the parameters of the van Genuchten (1980) equation based on different level of predictors' availability. The efficacy of the SVM approach was recently tested by Khlosi et al. (2016) for soils in the semiarid area of Syria. All mentioned authors have confirmed the outperformance of SVM to ANN techniques in terms of providing more accurate prediction of SWRC at specific points or parameters of SWRC models. Until now, this technique has not yet been applied to soils of the humid tropics.

1.3.2.3. k-Nearest Neighbor (kNN)

The k-Nearest Neighbor (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 space of multi-dimensional features (Nemes et al., 2006a). It is said to be 'lazy' since 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. Once the k-NN algorithm stores a set of training instances, application of the k-NN technique means identifying and retrieving the instances most similar to the target object from that set of stored instances, based on their input attributes (**Figure 1-3**). More theoretical details on this similarity based approach are given in Dasarathy (1991).

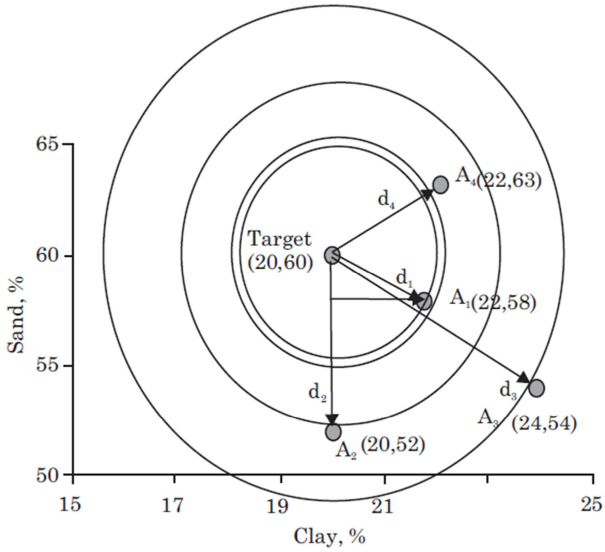

Figure 1-3. Graphical representation of k-Nearest Neighbors (kNN) for finding the best match for a target soil. Source: Jagtap et al. (2004).

The k-NN approach is considered by several authors (Bannayan and Hoogenboom, 2009; Buishand and Brandsma, 2001) as one of the most attractive pattern classification algorithms. Nemes et al. (1999) used a k-NN variant, to estimate missing PSD points from other existing PSD points to harmonize data of the European HYPRES database. 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. They also performed a detailed sensitivity analysis of this technique and found that the newly developed k-NN algorithm is robust in different scenarios. Based on the satisfactory results yielded by their k-NN algorithm, Nemes et al. (2008) introduced the user-friendly software called 'k-Nearest' that was developed with the option of estimating the uncertainty of the prediction. Elshorbagy et al. (2010a, b) identified the k-NN technique as an attractive modeling technique for hydrological applications because of its high level of flexibility. Recently, Patil et al. (2013) used the 'k-Nearest' software (Nemes et al., 2008) to estimate water content at -33 and -1500 kPa of 157 shrink-swell soils from India in order to derive their AWC. The ability of the k-NN approach to estimate water content at different matric potentials was tested and performed efficaciously to highly weathered soils in the humid tropics (Botula et al., 2013).

1.4. Type of soil water retention pedotransfer functions (SWRC-PTFs).

According to the review of Botula et al. (2014), soil water retention characteristic (SWRC) PTFs can be categorized into four general classification schemes (**Figure 1-4**). Depending on different criteria used by various authors, PTFs could be classified as (1) class PTFs and continuous PTFs; (2) point-based PTFs, parameter-based PTFs and pseudo-continuous PTFs; (3) Semi-physically based PTFs and empirically based PTFs; and (4) equation-based PTFs and pattern-recognition PTFs.

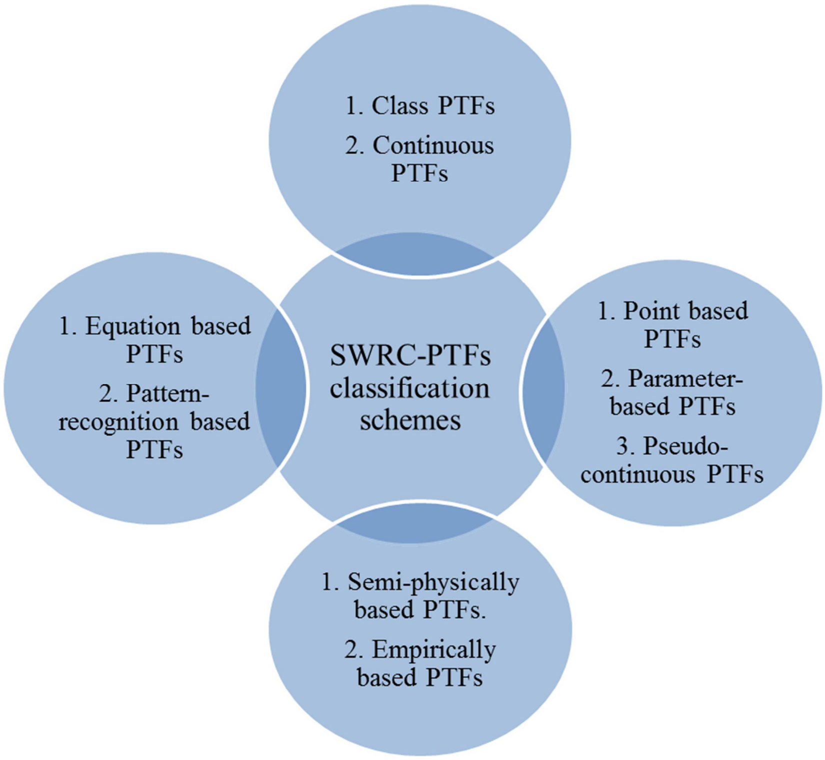

Figure 1-4. Four general categorization schemes of pedotransfer functions to estimate soil water retention characteristics (SWRC-PTFs).

1.4.1. Class and continuous PTFs

This classification scheme categorizes PTFs based on the availability of input information for SWRC estimation (Wösten et al., 1995; Wösten et al., 1990).

1.4.1.1. Class PTFs

Class PTFs provide an average estimation of soil hydraulic characteristics of specific soil classes. The development of class PTFs, also referred to as 'pedotransfer-rules', is based on preliminary grouping (Botula et al., 2014). Several grouping criteria have been commonly utilized to derive class PTFs, e.g. genetic-based groupings, horizon-based groupings, texture groupings, groupings based on soil structure and bulk density (BD), groupings of parent materials, and consecutive groupings of genetic horizons and then soil texture, or soil texture and then BD (Bruand, 2004a). For instance, McKeague et al. (1982) proposed class PTFs for the prediction of soil saturated hydraulic conductivity (Ks) based on morphological data associated with soil texture and soil structure characteristics. Batjes (1996) derived pedotransfer rules for the prediction of available water content based on FAO classification units together with soil texture and organic matter levels, whereas class PTFs for the prediction of field capacity (FC) and permanent wilting point (PWP) of 12 textural classes of US soils were introduced by Soil Survey Staff (1997) (**Table 1-2**).

In the literature, many class PTFs have been developed for soils of temperate regions, e.g. Carsel and Parrish (1988); Clapp and Hornberger (1978); Rawls et al. (1982); Vereecken et al. (1989), while it is rare for soils of the humid tropics. The shortness of class-PTFs for tropical soils is mainly due to the lack of large databases to provide a sound statistical-based grouping. The class PTFs of Hodnett and Tomasella (2002) developed for the parameters of the van Genuchten (1980) equation are among those few published for soils in the tropical regions.

1.4.1.2. Continuous PTFs

A continuous PTF is developed using the complete training database to derive regression equations (Wösten et al., 1990) in which the estimation of the hydraulic characteristics is based on continuous input variables, e.g., the real measured percentages of clay, silt, sand fractions; soil BD; and organic matter (OM) content (Wösten et al., 1995). Most existing PTFs developed to date fall in this category.

Because class PTFs provide only an average value of soil hydraulic properties for the whole soil class (e.g., texture class), developing and utilizing such PTFs are cheaper and easier than continuous PTFs. However, class PTFs generally seem to be less attractive than continuous PTFs due to their lower flexibility, limitation in providing site-specific information, and the occurrence of large errors in the estimation (Hodnett and Tomasella, 2002; Wösten and Nemes, 2004).

Obviously, Pachepsky and Rawls (2003) provided vivid illustration about the substantial error that one might encounter when using texture-based class PTFs developed from the US National Soil Characterization database. In **Table 1-2,** it can be obviously seen that although differences between average values of soil water retention of the textural classes are significant, the distribution of these values between classes overlap substantially. A similar uncertainty exists in texture-based class PTFs for estimation of saturated hydraulic conductivity also reported by Rawls et al. (1998).

Table 1-2. Water retention at two matric potentials in samples from different textural classes of US National Pedon Characterization database (Soil Survey Staff, 1997). Source: Pachepsky and Rawls (2003).

Textural class	Number of samples	Volumetric water content [a], m^3 m^{-3}	
		FC (at -33 kPa)	PWP (at -1500 kPa)
Sand	318	0.134 (0.072)	0.044 (0.025)
Loamy sand	528	0.172 (0.092)	0.062 (0.034)
Sandy loam	2984	0.238 (0.086)	0.096 (0.041)
Loam	2138	0.296 (0.067)	0.138 (0.042)
Silt loam	2791	0.334 (0.064)	0.132 (0.042)
Silt	51	0.335 (0.065)	0.085 (0.037)
Sandy clay loam	754	0.282 (0.062)	0.163 (0.038)
Clay loam	1203	0.345 (0.057)	0.203 (0.041)
Silty clay loam	1301	0.366 (0.047)	0.209 (0.037)
Sandy clay	141	0.301 (0.055)	0.209 (0.036)
Silty clay	661	0.403 (0.050)	0.268 (0.042)
Clay	1380	0.414 (0.068)	0.284 (0.049)

[a] Average values, with standard deviations in parentheses. FC is field capacity, PWP is permanent wilting point.

1.4.2. Point-based PTFs, parameter-based PTFs, and pseudo-continuous PTFs

Many researchers, e.g. Cornelis et al. (2001); Sharma et al. (2006); Wösten et al. (2001) among the others, make a distinction between PTFs that predict soil 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). Additionally, a recently published type of PTF that falls somewhere between the above two categories was introduced by Haghverdi et al. (2012) and was referred to as a pseudo-continuous PTF. **Figure 1-5** provides a schematic representation of point-based, parameter-based, and pseudo-continuous PTFs in case of using an artificial neural network (see section 1.3.2.1).

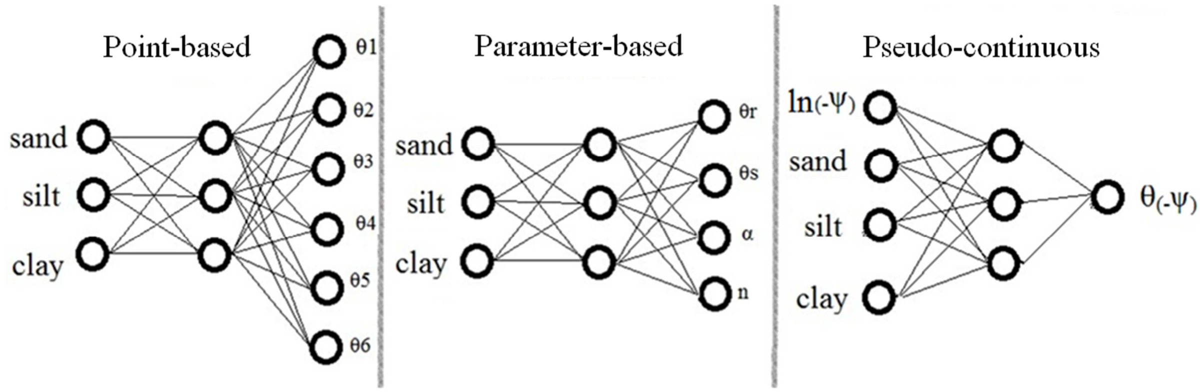

Figure 1-5. The typical topologies of the point-based, parameter-based and pseudo-continuous pedotranfer functions when using an artificial neural network (ANN). The first layer is the input layer having three input nodes/neurons (i.e., sand, silt and clay percentages). Bulk density, soil organic matter conntent, and other properties can be added as optional extra input neurons in other circumstances. The intermediate layer is the hidden layer whose number of nodes was determined by trial and error. The output nodes in the last layer were dertemined by the type of PTFs (e.g., for point PTFs, the nodes of θ_1, θ_2, . . ., θ_6 represent the volumetric water contents at specific matric potentials when using a data set containing six points of the water retention curve for each sample; whereas for parameter based PTF, θ_r, θ_s, α and n are are the outputs which in turn are the parameters of the van Genuchten equation (van Genuchten, 1980) in this example). For PC-PTFs, ln(-ψ) is matric potential which is the extra input predictor of the pseudo-continuous PTF. $\theta_{(-\psi)}$ is the volumetric water content at (-ψ) matric potential which is the output of the pseudo-continuous PTF. Source: Haghverdi et al. (2012).

The terminologies point-based PTFs and parameter-based PTFs used in this section are preferred to the more widely used point PTFs and parametric PTFs in the publications of Cornelis et al. (2001); Wösten et al. (2001) to avoid confusion with the term non-parametric PTFs found in the literature (Nemes et al., 2006a; Nemes et al., 2006b; Nemes et al., 2008).

1.4.2.1. Point-based PTFs

As we have mentioned above, point-PTFs provide the estimation of SWRC at particular matric potentials. For instances, Gupta & Larson (1979) used 43 different soil materials originating from 10 locations in the eastern and central USA to develop 12 PTFs for the estimation of soil moisture content at matric potentials ranging from -4 to -1500 kPa. Rawls & Brakensiek (1982) used a much larger database (N = 2543 horizons) collected across the USA to derive point PTFs for SWRC within the same matric potential range. Saxton et al. (1986) developed point-based PTFs from soils of the USDA database. Later, when a much larger USDA soil database became available (N=1722 samples), Saxton and Rawls (2006) reformulated the PTFs on the basis of those previously reported in Saxton et al. (1986) by including more variables and offering a wider range of applications. In this updating process, the initial equations were combined with equations of hydraulic conductivity, also considering the effects of soil bulk density, gravel, and salinity. These PTFs have been successfully applied to a wide variety of studies related to agricultural hydrology and water management, together with models like SPAW (Saxton and Willey, 2006) and AquaCrop (Steduto et al., 2008) through the user-friendly 'Soil Water Characteristics' software.

In the (sub)humid tropics, various efforts have been made to develop point-based PTFs from soil data sets specific to these regions. Most of these PTFs have been developed for application within restricted geographical domains for a limited range of soil textures and soil types. For example, Pidgeon (1972) derived point-based PTFs for the estimation of soil water content at field capacity (FC), permanent wilting point (PWP), and available water capacity (AWC) for ferralitic soils in Uganda. MacLean and Yager (1972) derived PTFs to predict AWC based on percentage of sand, silt, clay fractions, organic carbon (OC) content, and soil depth of soils in Zambia. FAO (1974) and Soil Survey Staff (1975, 1990) provided simple relationships between clay content and gravimetric water content at -1500 kPa (PWP) for ferralic and oxic horizons in various tropical regions. Lal (1978, 1981) derived point-based PTFs to predict gravimetric water content at -10 kPa, -33 kPa, and -1500 kPa and AWC based on a data set of soils developed from two different parent materials in Southern Nigeria. Recently, Minasny and Hartemink (2011) developed PTFs to predict water content at -10, -33, and -1500 kPa based on soil texture and BD of soils from the tropics. These soil data sets are parts of the IGBP-DIS (International Geosphere Biosphere Programme Data and Information System) soil database obtained from ISRIC (International Soil Reference and Information Center) in Wageningen (the Netherlands). Chakraborty et al. (2011) developed PTFs from a wide textural range of Indian soils. Obalum and Obi (2012) proposed point-based PTFs for kaolinitic and coarse-textured tropical soils from southeastern Nigeria. Santos et al. (2013) generated and validated PTFs to predict gravimetric water content at -33 and -1500 kPa for different soil classes from the central-south of the State of Rio Grande do Sul in Brazil.

1.4.2.2. Parameter-based PTFs

Parameter-based PTFs are predictive models that provide the estimates of parameters of analytical expressions describing the SWRC, such as the Brooks and Corey (1964), the Campbell (1974), and the van Genuchten (1980) equations. Parameter-based PTFs generate continuous curves describing the hydraulic characteristics of soils which in turns is very important for modeling purposes (Cornelis et al., 2001; van den Berg et al., 1997), and allow the computation of hydraulic values at arbitrary pressures (Børgesen and Schaap, 2005). The first parameter-based PTFs were developed using data sets of soils from temperate regions, e.g., Cosby et al. (1984) and Rawls and Brakensiek (1985) developed regression equations for the Brooks and Corey (1964) model based on soils from the USA. Saxton et al. (1986) used the percentage of clay and sand to calculate the parameters of a model derived by Campbell (1974). Vereecken et al. (1989) developed widely used PTFs for estimation of the parameters of the van Genuchten (1980) model based on sand, clay, OC, and BD of 182 horizons of 40 different soil series in Belgium. Wösten et al. (1999) predicted the parameters of the van Genuchten (1980) model using the HYPRES database including data of 4030 horizons from all over Europe.

For tropical regions, parameter-based PTFs were developed by van den Berg et al. (1997) to predict the soil water retention parameters of the van Genuchten

(1980) equation. Tomasella and Hodnett (1998) developed PTFs to predict the parameters of the Brooks and Corey (1964) equation from soil texture and OC using a data set of various soils from Brazilian Amazonia. Tomasella et al. (2000) derived parameter-based PTFs for the van Genuchten (1980) equation, although they experienced that the van Genuchten (1980) analytical function may not be the best one to properly describe the hydraulic behavior of Oxisols. Earlier, van den Berg et al. (1997) found that the van Genuchten (1980) equation can adequately describe SWRC of soils with low activity clays in the southern part of Brazil. Later, Hodnett and Tomasella (2002) arrived at the same conclusion for Brazilian soils. Hodnett and Tomasella (2002) used 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 data set as the IGBP/T data set, which exclusively contained soils from tropical climates. Santra and Das (2008) developed parameter-based PTFs for the van Genuchten (1980) model to predict SWRC of soils from a hilly watershed in eastern India. Adhikary et al. (2008) did the same for the Brooks and Corey (1964) model to provide the prediction based on soils from various parts of India.

In their recent review paper, Botula and co-workers (2014) summarized that most of the PTFs developed for soils in the (sub)humid tropics were point-based PTFs. The possible explanation of such bias is that point-based PTFs statistically outperform the parameter-based PTFs (Dashtaki et al., 2010; Pachepsky et al., 1996b; Tomasella et al., 2003; Vereecken et al., 2010). The well performance of point-based PTFs is in part attributed to the fact that water content is controlled by different soil properties, depending on the level of soil matric potentials. Hence, point-based PTFs allow more appropriate independent variables to describe the water content variation than do the parameter-based PTFs. However, most point-based PTFs are often limited to the prediction of water content at certain matric potentials, generally field capacity (FC) (i.e., at -10 and -33 kPa) and permanent wilting point (PWP) (i.e. at -1500 kPa). These values are typically used to calculate the water depth that should be applied through 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, as several physically based models (e.g., HYDRUS, SWAP, SWATRER, MACRO, etc.) require the information of a complete SWRC as input, the PTFs that provide continuous estimation of SWRC are therefore still in great demand.

1.4.2.3. Pseudo-continuous based PTFs (PC-PTFs)

The use of parameter-based PTFs exposes several drawbacks in describing the whole SWRC (e.g., the real shape of SWRC is not always identical to the one of the selected equation, and the SWRC formed by outputs of parametric PTFs always carries more error than the fitted SWRC) (Haghverdi et al., 2014). A new PTF approach, named ‘pseudo-continuous PTFs’ (PC-PTFs), was therefore recently introduced by Haghverdi et al. (2012) with an expectation of filling the gap between the need of continuous estimation of SWRC and the large uncertainty in parameter-based PTFs. PC-PTFs use the natural logarithm of matric potential as an input parameter which in turn enables the prediction of water content at any desired matric potential. There is only one output parameter, $\theta(-\psi)$, which shows the water content at the predefined matric potential, ψ. Different values of matric potential yield different water contents. Haghverdi et al. (2012) proved that PC-PTFs derived by the Artificial Neural Networks (kNN) technique were more accurate and reliable than parameter-based PTFs, and slightly better than point-PTFs when a limited data set was available for PTFs’ development. This recent approach has only been tested for soils of temperate and (semi)arid regions using ANN and Support Vector Machines (SVM) techniques (Haghverdi et al., 2012; Haghverdi et al., 2014).

1.4.3. Semi-physically based PTFs and empirically-based PTFs

According to McBratney et al. (2011), there are two major approaches to derive PTFs: a semi-physical, which attempts to describe a physical or chemical model relating the basic properties to the predicted properties; and an empirical approach, which is the most widespread, linking the basic soil properties to the more difficult-to-measure soil properties by means of fitting regression models to observed data.

1.4.3.1. Semi-physically based PTFs

Semi-physical methods recognize the similarity between the shape of the particle size distribution (PSD) and SWRCs. They offer valuable conceptual insights into the physical relations between PSD and pore size distribution (POD). 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 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 the degree of irregularity 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 to all of it. 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).

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 data sets and two unpublished data sets 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 a 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.

The main disadvantage of these methods is that they often require a very detailed PSD, making them almost as difficult to apply as direct measurements (Schaap, 2005). Moreover, because the pore-solid relationships are affected not only by the distribution of solid particles size but also by their spatial arrangement (i.e., soil structure), using only detailed PSD may not be sufficient to predict SWRC in the wet range (Giménez et al., 2001). Significant errors of using physical-conceptual PTFs were recorded in many case studies (Cornelis et al., 2001).

1.4.3.2. Empirically-based PTFs

The empirical approach is the one that is most used to develop soil water retention PTFs in both temperate and tropical regions. The most commonly used techniques for fitting or deriving PTFs are statistical regressions, Multiple Linear Regressions (MLR) and polynomials of the n^{th} order. Other modern numerical and statistical methods applied are Generalized Linear Models (GLM), General Additive Models (GAM), the Group Method of Data Handling (GMDH), and Multiple Adaptive Regression Splines (MARS). Currently, data-mining techniques are gaining popularity in the PTF-research field with the application of nonconventional statistical methods, e.g., Artificial Neural Networks (ANNs), Classification and Regression Trees (CART), k-Nearest Neighbors (k-NN), Support Vector Machines (SVM), Genetic Algorithms (GA), and Genetic Programming (GP). Some of the widely applied regression methods to derive empirically-based PTFs were described in section 1.3

1.4.4. Equation-based PTFs and pattern-recognition PTFs

The PTFs described above can also be categorized as 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 to some extent. In contrast, in pattern-recognition PTFs, a priori model does not need to be defined. They are based on pattern recognition and make use of the recently developed data-mining and machine-learning techniques: ANN, Regression Trees (RT), k-NN, SVM, GP. Detail description of commonly used pattern-recognition methods was presented in section 1.3.2.

1.5. Objectives of the study

Although considerable progress has been made in developing hydraulic PTFs for tropical soils (Botula et al., 2014), several persisting regional knowledge gaps about soil hydraulic properties in the tropics has so far not been addressed. This issue was recently confirmed by Pachepsky et al. (2015) in their review on PTF development and utilization. These authors noted that remarkably little effort was put into PTF development for saline soils, calcareous and gypsiferous soils, peat soils, and paddy soils. Indeed, as far as we know, no SWRC-PTFs have been developed for tropical delta soils where the main agricultural practice is paddy rice cultivation and where the soil-water relationship is not well understood.

Therefore, the main objective of this dissertation is to better understand the complex soil-water relationships of soils in the tropical Vietnamese Mekong Delta by specifically examining how pedological attributes affect soil water retention through the development of innovative ‘hydraulic’ pedotransfer functions. The information of soil hydraulic properties is then a prerequisite input of decision support tools for sustainable agricultural management in Vietnamese Mekong Delta.

To address the main study objective, five specific objectives (SO) were formulated and reflected in chapters 3, 4, 5, 6, and 7 of the dissertation.

1. Since developing new PTFs is a very arduous task which requires a large soil database of good quality, utilizing existing PTFs in the literature is wise, though their validity should be tested first. The first objective of this dissertation was therefore to evaluate the performance of several well-known published PTFs derived from both ‘temperate’ and ‘tropical’ climates in predicting SWRC of tropical Vietnamese Mekong Delta soils.

2. With an attempt to understand how pedo-genesis properties of tropical Vietnamese Mekong Delta soils (e.g., soil texture, soil structure, field morphological data, other physic-chemical soil properties as BD, OM or OC content, etc.) relate to soil hydraulic characteristic (i.e., SWRC), innovative PTFs were developed for Vietnamese Mekong Delta soils based on the information of a limited local data set (N=160 observations). Several strategies in PTF development, e.g. regression methods and potential predictors, were taken into consideration to obtain the most accurate and reliable PTFs. The second objective of the study was to investigate how closely readily available soil parameters, particularly categorical soil structure information, available in local data sets, are related to SWRC?

3. The third objective was to examine the predictive power of models derived by different regression methods. To that end, statistical regression and three data-mining techniques were evaluated in terms of their accuracy in predicting SWRC.

4. The fourth objective was to address the combined effects of incorporating categorical soil structure information into different regression methods on derived PTFs’ accuracy.

5. The final objective was to assess the utility of the PTFs that were derived and validated in this dissertation through functional validation. Such evaluation was conducted with the aim of examining the validity of the derived PTFs with respect to specific practical uses, e.g., to simulate yield crop response to soil water and irrigation requirement using the crop-water model AquaCrop, or to investigate the variation of soil water content and degree of water stress using the hydrological model Hydrus-1D. To that end, the effect of replacing laboratory-measured SWRC data with predicted data obtained from locally-derived PTFs and globally-offered PTFs embedded in soil water models (e.g., point PTFs of Saxton and Rawls (2006) and parameter-based neural networks PTFs of Schaap et al. (2001)) on the variation in the simulated outcomes of the models was investigated.

1.6. Outline of the thesis

The dissertation is constructed based on the general and specific objectives outlined above and are divided into eight chapters. After the introductory Chapter, Chapter 2 gives a general description of the bio-physical and social-economic settings of the study area, as well as the locally collected data used for pedotranfer modeling. Chapter 3 presents the evaluation of the applicability and reliability of published PTFs in predicting soil water retention characteristics of tropical Vietnamese Mekong Delta soils. Chapters 4, 5, and 6 deal with some important strategies in PTF development in order to derive accurate and reliable PTFs. Specifically, in Chapter 4, potential predictors for SWRC estimation of Vietnamese Mekong Delta soils were investigated using a multiple linear regression approach, in which specific attention is given to the use of categorical soil structure information which is normally available in many soil survey databases to improve the accuracy of derived PTFs. Chapter 5 compares the performance, in terms of accuracy and reliability, of point and pseudo-continuous PTFs developed by different regression techniques, such as MLR, ANN, SVM, and kNN. Chapter 6 investigates the combined effect of the two improved strategies presented in Chapters 4 and 5 in order to get the best predictive models. Chapter 7 presents the functional evaluation of derived PTFs by monitoring simulated soil water balances using agro-hydrological models (i.e., AquaCrop and Hydrus-1D). Finally, Chapter 8 summarizes the key findings of the study and gives some recommendation for future research.

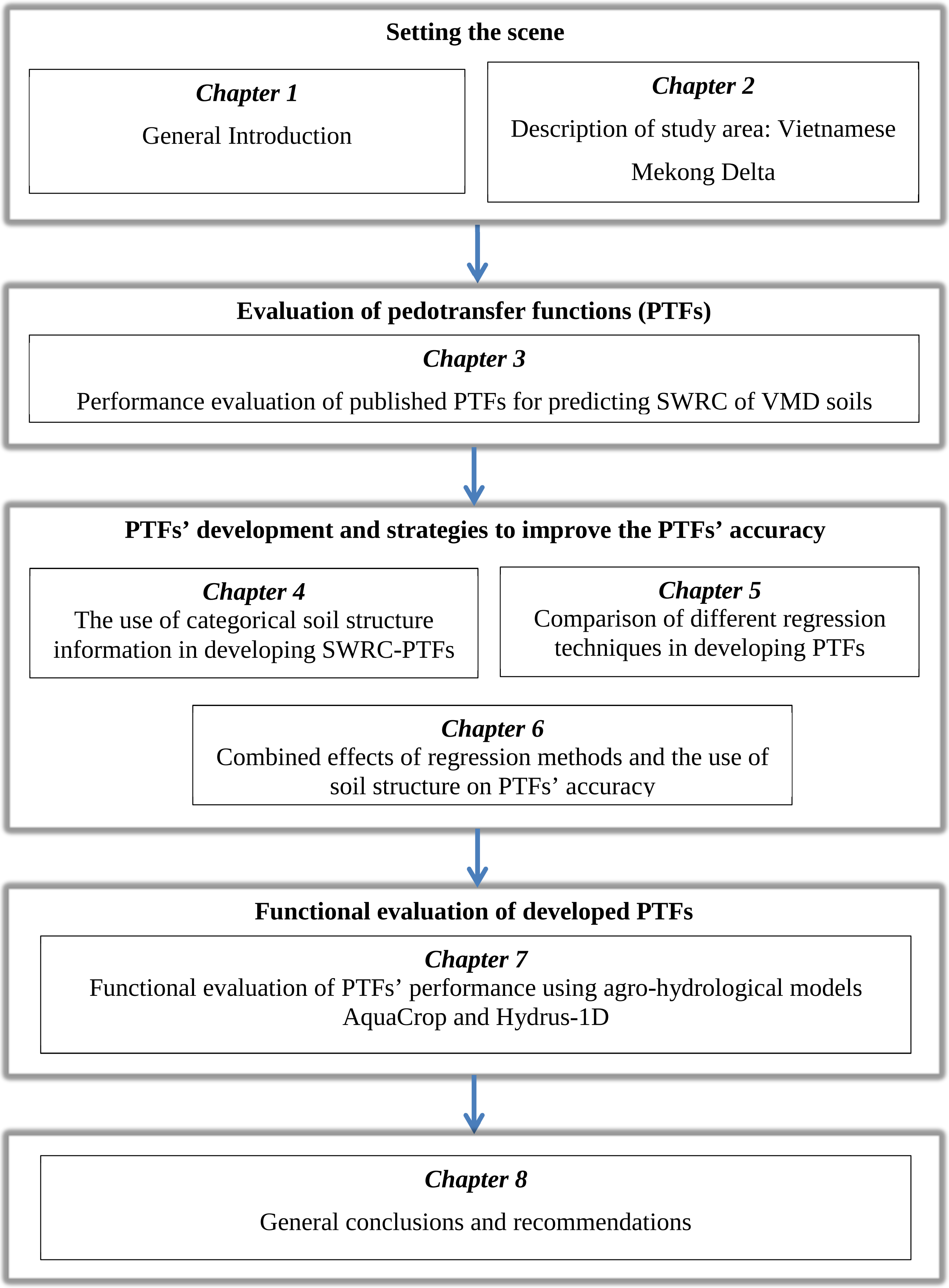

Figure 1-6. Flow-chart of the

Chapter 2

DESCRIPTION OF THE STUDY AREA

2.1. Biophysical characteristics

2.1.1. Geographical location

The study area, the Vietnamese Mekong Delta (VMD), is located in the Southernmost part of Vietnam and at the downstream end of the Mekong River. This wide and flat delta is characterized by perpetual sedimentation at the latter's mouths and covers an area of 4 million hectare (approximately 79% of the total area formed by Mekong River and 12% of total natural land of Vietnam). It spreads from 11^{o}N to $8^{o}30'$N latitude with the three sides connected to the sea and another border jointed to the mainland (Tran et al., 2007) (**Figure 2-1**).

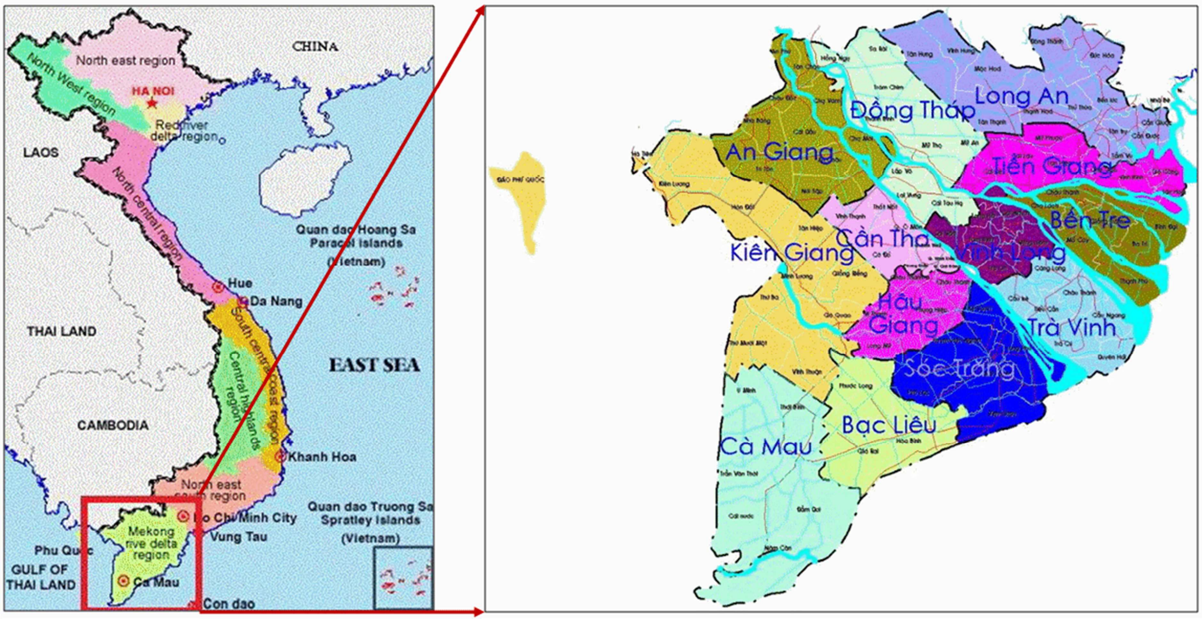

Figure 2-1. The possition of the Mekong Delta in Vietnam and it's corresponding admistrative map.

2.1.2. Climate and hydrology

The whole delta is influenced by a tropical monsoon climate with two typical distinguished seasons: a wet season (summer monsoon with South-Western wind direction) from May to November with high precipitation (about 90-94% of the total rainfall over the year), and a dry season (winter monsoon with North-Eastern wind direction) from December to April. Average annual rainfall in the delta ranges from 1400 mm (± 350 mm) in the central to 2400 mm (± 240 mm) in the western part, while the eastern part is receiving on average 1600 mm (± 220 mm) of rainfall. There are about 107-165 rainy days in a year. Temperature in the delta is high and stable, averagely ranging from 26^{o}C to 28^{o}C across the entire area. Evaporation is about

1100 – 1400 mm depending on the specific topography of the area. The relative humidity ranges from 79% to 85% (Le, 2003).

The hydrological regime of the Mekong Delta is controlled by the Mekong River's discharge and Sea's tide. The upstream delta is mainly influenced by Mekong river. The main Mekong River is rather long (about 4200 km), it terminates and branches into two main tributaries when it enters into the Mekong delta (i.e., Tien and Hau Rivers of about 250 km long). The highest discharge of the rivers is in the flood period of August – October, and the lowest one is in the dry season (March – May). The high annual rainfall combined with the high level of the Mekong River results in regular floods of 0.3 to 3 m during the wet season (mainly from July to November, with highest flood level in September). The downstream delta, on the other hand, is mainly affected by diurnal tidal movements of the Eastern Sea and the semi-diurnal tidal movement of the Western Sea (i.e. the Gulf of Thailand). Due to the low-lying position together with the combined effect of the lowest discharge of Mekong river within the delta and the sea tidal regimes in the dry season, saline water can encroach into the delta for about 40-50 km.

Generally, climatic and hydrological conditions of VMD have strong impacts on the soil forming processes as well as the spatial and temporal variability of VMD soils (Le, 2003). For instance, in the dry season, the rainfall is small (10% of the annual rainfall), the river discharge is low, and evaporation is rather high. These conditions together cause a severe water imbalance (Estellès et al., 2002) and promote inland salinity intrusion and the change of potential acid sulfate soils into actual ones (Vo and Nguyen, 2012). In the rainy season with water surplus, salts and sulfuric acids are leached making these problem soils suitable for agricultural production.

2.1.3. Geology

The Mekong Delta is a typical peninsular land of Vietnam. It is a young delta deposited by the Mekong river network, not more than 10,000 years ago. About 2 million years ago, the whole South-East Asia was lifted up above sea level and an old alluvial sedimentation plain was formed, the so-called Pleistocene delta consisting of coarse materials (Driessen and Dudal, 1989). Most of the delta, at the end of the Pleistocene period, was a mass of muddy land and mainly impacted by encroachment and erosion processes. The sea level was low at that time, about 120 m lower than at present. Up to the early Holocene period, the sea progressed gradually into the low-lying delta and formed a shallow sea type. In the middle of the Holocene period, the sea again rose greatly over the area and the delta was deeply submerged under sea level. Sea progression and regression produced a rather thick sedimentary layer. Eventually, when the sea level gradually withdrew from the delta, sediment from the Mekong River was deposited and the current Mekong Delta was formed during the end of Holocene period (Tran, 1998).

The present delta consists of levees, floodplains and sandbars parallel to the coastline. Between the sandbars, natural vegetation (i.e., mangrove forests) has added organic matter to both fluvial and marine sediments, providing ideal conditions for the formation of acid sulfate soils. The combined action of the river and the sea has formed good alluvial soils along the river and acid sulfate soils in the back-swamps.

According to Mensvoort and Van den Berg (1986), six geological formations are distinguished and distributed as shown on the map in **Figure 2-2**: (1) Peaty dome, (2) Holocene brackish water and marine deposits in depression area, (3) Pleistocene alluvial deposits, (4) Granite hills and mountainous area, (5) Holocene fluvial deposits, complex of levees and back-swamps, and (6) Beach ridges and clay plains.

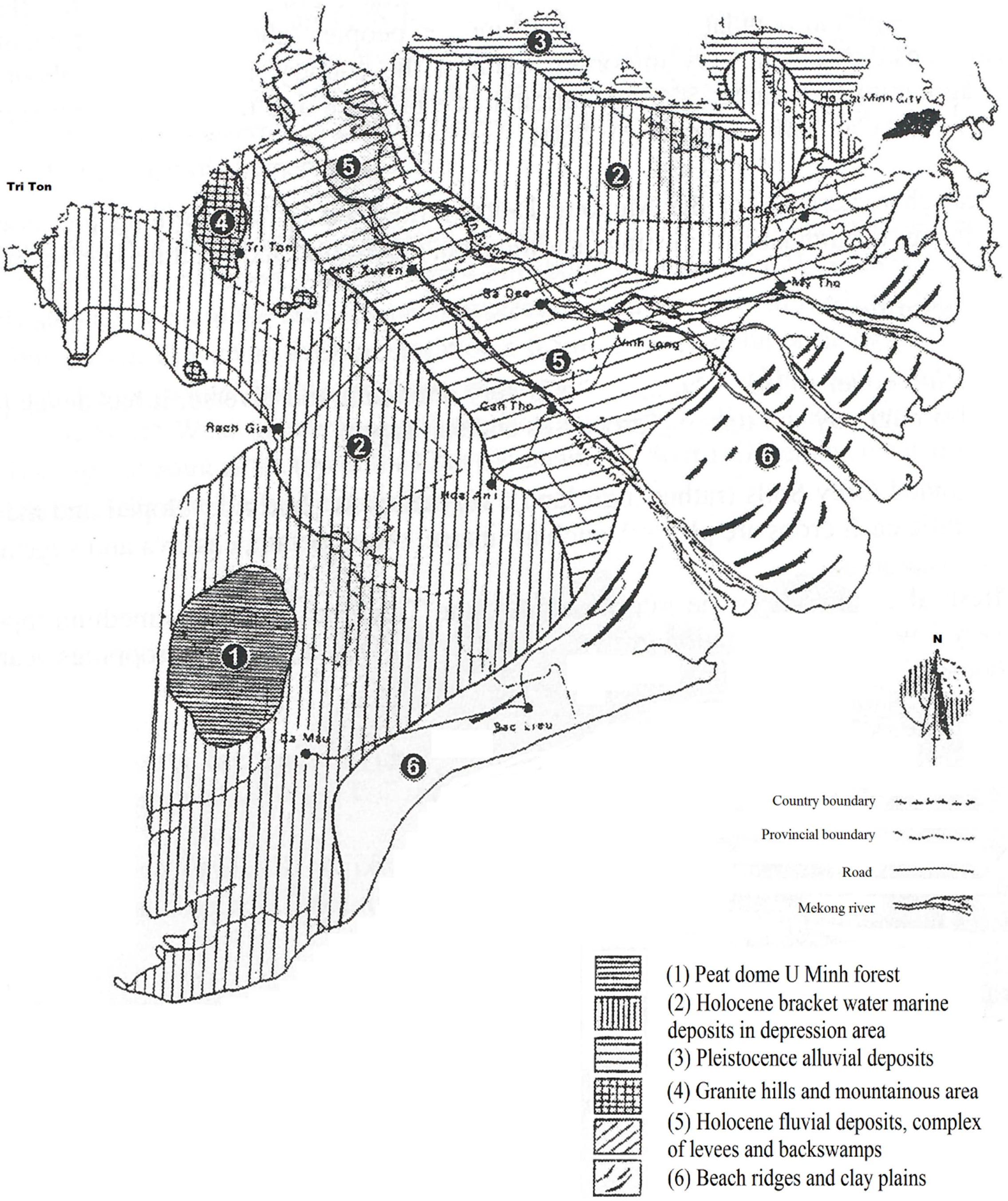

Figure 2-2. Map of geology and physiography of the mainland of the Mekong Delta in Vietnam simplified after Mensvoort and Van den Berg (1986). Source: Le (2003)

2.1.4. Vegetation and land use type

The vegetation cover in the Mekong Delta is very abundant and diverse. It is influenced by soil forming processes, soil types, and the availability of water resources. For instance:

- In the upland region with degraded grey soils (see section 2), vegetation is poorly developed and mainly cash crops are cultivated (i.e., maize, sweet potato, cassava and vegetables).

- In the region with fresh alluvial soils located in the interim area of the Tien and Hau Rivers (medium topography), very little natural vegetation is present and the main vegetation found in this area are rice and fruits from intensive paddy rice fields and fruit orchards nearby the farmers' houses.

- In the lowland area with acid sulfate soils, many indicative natural grass species (especially *Eleocharis dulcis* and *Scleria poaeformis*) and wild forests (*Melaleuca sp.*, *Eucalyptus sp.*) are present.

- In the regions with alluvial soils with temporal intrusion of saline water, many kinds of salt-tolerant vegetation are present and one to two rice crops are cultivated per year depending on the availability of fresh water; alternatively, shrimp-rice cultivation is also popular in this region.

- In the coastal region with saline soils, salt-tolerant grasses and tropical mangrove forests are the prominent vegetation.

2.2. Socio-economic situation

The Vietnamese Mekong Delta (VMD) is the most populated area of the Mekong River basin with a population density of 424 persons per km^2. The region is home to over 17 million inhabitants (about 20% of Vietnam's population). Only around 25% of the population is urbanized, leaving some 75% of population rural. The livelihood, cultures and economies of major part of VMD's people are related to agricultural production which in turns is closely associated to the Mekong river.

The Mekong Delta plays a very important role in the social-economic development strategy of Vietnam and is the main key for food security of the country. Its economy was, is, and will continue to be dominated by agriculture, mainly paddy rice production and the fishing industry. To date, approximately 2.6 million ha, around 63% of the total land area in the VMD, is used for rice production, compared with 28% for whole Vietnam and 38% for the Red River delta. The delta accounts for 53% of the nation's rice production and more than 80% of the rice is used for national export (GSO, 2009) positioning Vietnam as the 5th largest producer of paddy rice and the 2nd largest exporter of milled rice worldwide (FAOSTAT, 2012)

Together with the favorable climatic conditions as well as the well-established canal systems for irrigation in the dry season, the fertile soil resource of VMD enables highly productive agriculture in the region. However, since paddy rice culture and fishing industry are extensive users of fresh water, agricultural economy of the delta totally depends on and is tremendously controlled by natural resources (i.e., soil and water). There are several natural constraints which strongly affect agricultural production in the region, such as: flooding in the wet season; acid sulphate soils, and their effects on soil productivity, drainage water quality and aquatic productivity; water shortage and saline intrusion in the dry season; depletion of coastal mangroves and protected areas for fish breeding.

Furthermore, the effect of climate change on the socio-economy of the delta is also important to notice. Indeed, climate is already changing in the delta and the Mekong Delta is ranked amongst the top five deltas in the world most likely to be severely affected by climate change. According to the assessment report of the Intergovernmental Panel on Climate Change (IPCC, 2007), trends in temperature, rainfall and sea level are noticeable with a limited record of meteorological and hydrological data in the Mekong Delta. From 1970 to 2007 the average temperature rose with 0.6℃ and rainfall increased with 94 mm. Predictions of climate change in Vietnam in the period to 2050 which were carried out by the Vietnam National Institute of Meteorology, Hydrology and Environment (IMHEN) have showed that rainfall tends to decrease in the dry season and to increase in the wet season (Mekong Delta Plan, 2013). Together with the temperature rise and changing rainfall patterns, sea level rise is also expected to have a huge impact on the physical conditions of the Mekong Delta, leading to a range of effects on people, their health, livelihood and prosperity. In the flat areas of the delta, the predicted sea level rise can result in large areas of permanent and more frequently inundated coastal plains. Depending on the scenario, the percentage of inundated delta ranges from 12.8 to 37.8%. Rice production will be affected through excessive flooding in the tidally inundated areas and longer flood periods in the central part of the delta. These adverse impacts could affect all three cropping seasons.

2.3. Soils

2.3.1. Previous pedological studies about VMD soils

A large number of pedological studies about VMD soils has been conducted at the Department of Soil Science of Can Tho University. Within the period 1980-2006, extensive studies, funded by many international and interuniversity cooperation projects (e.g., VH-10 and MHO-8 projects with Wageningen University for the integrated managements of acid sulfate soils and soils in the coastal area, SAREC project sponsored by SIDA-Sweden for VMD problem soils, VLIR projects with Flemish Universities of Belgium for studying soil dynamics in the terrestrial environment, and many others), were carried out to investigate and characterize soil resources in the Mekong Delta.

Regionally, VMD soils were classified into five general major soils groups based on soil fertility capacity for rice production (Vo, 2006). They include:

- Alluvial soils which dominate along the Bassac and Mekong Rivers and cover an area of 1.1 million ha (30% of the delta). These soils are considered as the most productive soils in the Delta with the average of 2% organic matter content, a total N range of 0.1 to 0.25%, and a medium level of Phosphorus and Potassium content (Estellès et al., 2002). Two to three rice crops can be grown on these soils each year.

- Acid sulphate soils (ASS) which cover an area of 1.6 million ha (41% of the delta) and are found in depression (back-swamps) regions where there is no or little cover of river sediment deposits. This soil group can be further divided into different subgroups, such as actual and potential ASS with and without saline intrusion, and peaty ASS. These soils are rich in organic matter and total N, low in P and high in soluble Al and Fe. The main limitation of this soil group to crop production is the high level of phytotoxic substances.

- Saline soils which occupy about 20% of the total area (0.7 million ha) and occur along the coastal area from the east to the west. This group of soils can be divided into two subgroups: temporary dry-season saline soils and permanent saline soils. These soils are rich in nutrients but the high level of salinity limits plant growth.

- Sand-ridge soils which are distributed along the eastern coast (43,000 ha or 1.5 % of the total area). Rice and upland crops are cultivated on these soils. The position on a sandbar or between sandbars dictates the texture of the soil.

- Degraded grey soils which occupy only a small area in the northern part of the delta (150,000 ha or 4% of total area), but play an important role in some provinces (e.g., An Giang and Long An) as the main soil resource in these areas. They can be divided into two groups: soil developed on acidic magmatic rocks and soils developed on old alluvial deposits. This soil group is not favorable for crop production as nutrients are depleted and soil microbial activity is generally low in these soils.

Following the USDA/Soil Taxonomy and FAO/UNESCO classification systems, the soils in VMD are classified into major soil groups as presented in **Table 2-1**. The distribution of different soil groups in the region was illustrated in **Figure 2-3**.

Table 2-1. Major soil groups in the Vietnamese Mekong Delta.

Soil group	USDA-Soil Taxonomy	FAO/UNESCO	Area (ha)
1. Alluvial soils			1,184,857
Tidal flat alluvial soil	Fluvaquents, Ustifluvents	Fluvisols	
Developed alluvial soils	Ustropepts, Tropaquepts	Gleysols	
2. Acid sulfate soils (ASS)			1,687,000
Slight potential ASS	Sulfaquents	Sulfi-Thionic Fluvisols	
Moderate to severe potential ASS	Sulfic Tropaquents	Sulfi-Thionic Fluvisols	
Severe actual ASS	Sulfaquepts	Orthi-Thionic Gleysols	
Moderate actual ASS	Sulfic Tropaquepts	Orthi-Thionic Gleysols	
Saline-slight potential ASS	Salic Sulfaquents	Sali-Sulfi-Thionic Fluvisols	
Saline-moderate potential ASS	Salic-Sulfic Tropaquents	Sali-Sulfi-Thionic Fluvisols	
Saline severe actual ASS	Salic Sulfaquepts	Sali-Orthi-Thionic Gleysols	
Saline moderate actual ASS	Salic-Sulfic Tropaquepts	Sali-Orthi-Thionic Gleysols	
Peaty acid sulfate soils	Sulfihemist	Thionic Histosols	
3. Saline soils			682,262
Permanently saline soils	Salic Hydraquents	Gleyic Solonchaks	
Severe saline soils	Salic Fluvaquents, Salic Ustifluvents	Stagni Salic Fluvisols	
Slight to moderate saline soils	Salic Tropaquepts, Salic Ustropepts	Stagni Salic Gleysols	
4. Degraded grey soils and others	Tropaquults, Plinthaquults	Acrisols	145,763
5. Sandy soils	Fluventic Tropapsamments	Haplic Arenosols	43,318

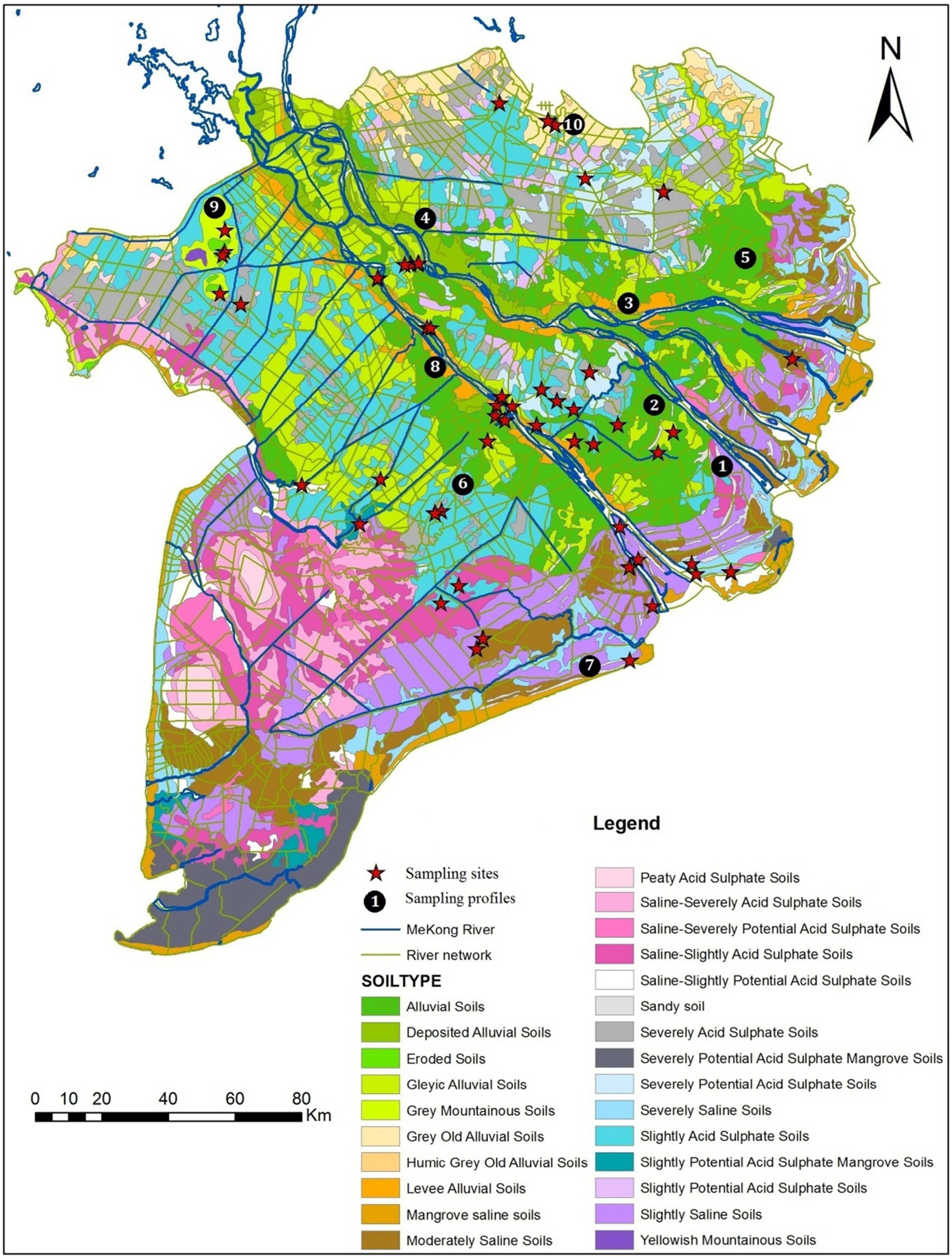

Figure 2-3. The map of different soil groups in the Vietnamese Mekong Delta. Red stars represent the sampling sites of traing data set (i.e., data collected in the framework of this dissertation). Circles with number display the location of soil profiles of test data collected from the study of Le (2003).

Regarding mineralogy of VMD soils, Le (2003) found that quartz is the dominant mineral over others such as feldspars, kaolinite and chlorite in the silt fraction. In the clay fraction of recent alluvial soils, muscovite is the predominant mineral beside smectite and vermiculite, whereas kaolinite (1:1 clay minerals) is prevailing in well-developed Pleistocene alluvial soils and weathered soils in mountainous areas from sandy and colluvial materials.

Moreover, due to the recent origin of the soils form in the young river delta basin, the majority of soils in the study regions shows weak structural development (Le, 2003). This aspect has a strong influence on soil hydro-physical behaviors, particularly on agricultural soils, as these soil properties have been strongly influenced by agricultural activities (e.g., pudding to destroy soil structure for rice cultivation, or raising beds to cultivate upland crops, etc.).

2.3.2. Soil samples used for pedotransfer modeling

Although a large volume of morphological and physico-chemical soil information is available in many soil survey databases as well as in data sets from various soil fertility experiments, information on soil hydraulic properties in VMD is very rare and not sufficient for PTF development.

In the framework of this study, two local data sets taken from samples of the study area were utilized for evaluation of published PTFs (**Chapter 3**), development of new SWRC-PTFs for tropical VMD soils (**Chapter 4, 5, 6**) and validation of their performance in both statistical and functional behaviors (**Chapter 5 and 7**). The information of sampling sites, methods of soil analysis, and statistical distribution of soil properties in two data sets was described in detail as below.

2.3.2.1. PTFs' evaluation and calibration data set

The first data set, which was used to evaluate the published PTFs (**Chapter 3**) and to develop new PTFs for VMD soils (**Chapter 4, 5, 6**) was constructed in the framework of this study through a local field campaign conducting in the period of August 2010 to January 2011. Since our main objective was to construct PTFs for estimating SWRC in a context of agricultural water management, the training data set were collected with the aim of covering a wide range of soils primarily exploited for agricultural production in the VMD (mainly paddy rice, but also upland crops such as vegetable, maize, and sugar cane).

As presented in the study of Le (2003), the soil profile description of different soil types displayed similar distribution of master horizons throughout the profiles (e.g., Bg or Bgj master horizons underlaid the plough Ap surface horizon). Similar patterns of horizon distribution were observed again in this soil campaign (by using one meter auger for soil profile investigation). For agricultural production, the first two master horizons are of significant importance for crop development in terms of providing the base for root to grow and to take-up water and nutrients.

Considering that the spatial variation of soils over the landscape might be much more important in explaining the variability of SWRC in the study region than a vertical change within the soil profile, we only took the samples from the two first master horizons (instead of the whole soil profile). More consideration was taken to spatial variability by increasing the sites that the samples were taken in horizontal direction. The locations of sampling sites were indicated in **Figure 2-3**.

During the soil survey, 160 soil samples (both disturbed and undisturbed samples) were taken from two upper diagnostic horizons in agricultural fields with different land use types. The depth of the two upper diagnostic horizons varies from site to site, with a maximum lower boundary at a depth of 25 cm for surface (Ap) horizons and 70 cm for subsurface horizons (Bg, Bj, or Bgj).

The SWRC as well as other chemical, physical and morphological properties were quantified by in-situ field and ex-situ laboratory techniques. Standard methods, described in e.g., Page (1982) and Dane and Topp (2002), were used to determine all variables in the laboratory. In brief, the undisturbed soil samples (with 4 replications), taken by standard Kopecky rings of 100 cm^3 in volume, were used to determine soil bulk density (BD) by the core method (Grossman and Reinsch, 2002), and SWRC at eight matric potentials (e.g., -1, -3, -6, -10, -20, -33, -100, -1500 kPa) using sand-boxes and pressure chambers according to the procedures outlined in Cornelis et al. (2005).

The disturbed soil samples taken nearby the place of undisturbed sampling pits were utilized to determine other chemical and physical soil properties. Specifically, dry 2 mm sieved soil samples were used to determined soil organic carbon content by wet combustion method (Walkley and Black, 1934), pH_{H20} and EC both at the 1:2.5 ratio of dilution, particle size distribution by sieve-pipette method

(Gee and Bauder, 1986), soil aggregate stability by dry and wet sieving method of De Leenheer and De Boodt (1959) and soil plastic limit (ASTM, 2010). Soil plastic limit was determined as the gravimetric water content at which a soil sample could be rolled by hand into a thread of 3.2 mm diameter without breaking.

Soil morphological characteristics (soil color at sampling and at dry state, soil horizon, depth of soil horizon, soil structure in terms of presence and absence of pedality, grade of structure development, types and size of structured units) were described directly in the field during the soil survey or afterwards in the laboratory according to the FAO Guidelines for Soil Description (FAO, 2006). Supplemental information of soil groups, geological units, diagnostic horizons, and depth of horizons of each soil samples in the data set can be referred in the Appendices.

2.3.2.2. PTFs' validation data set

The performance of PTFs derived in **Chapter 5** was validated by an independent data set. The validation data set used in this study was obtained from the study of Le (2003). This set includes complete records of 29 samples taken from 10 soil profiles which are representative for several major soil groups within the study region. The physical and chemical soil properties, and SWRC of the test samples were determined by the same methods as mentioned above for the training data set. Originally, the SWRC of test samples were determined at nine matric potentials (i.e. -0.25, -1, -3, -5, -7, -10, -33, -100, -1500 kPa). The eight soil water retention points corresponding to those of the training samples were obtained based on fitted van-Genuchten parameters (van Genuchten, 1980). Additional information of soil and geology of 10 soil profiles in the validation data set was presented in the Appendices.

2.3.2.3. Soil properties of the data sets used in this study

Descriptive statistics of the basic soil physical, chemical and SWRC of two data sets are summarized in **Table 2-2**. The soils in the study area have a wide ranges of basic soil properties, e.g., the range of sand, silt, clay content, bulk density and organic carbon content is 0.13–98.6%, 0.00–64.9%, 1.4–76.8%, 0.7–1.9 Mg m^{-3}, and 0.08–12.3 %, respectively for the training data set. The testing dataset has similar ranges, except for OC content. The wide ranges in soil properties is primarily associated with the alluvial soils in the study. The soils can be assumed as medium-to fine-textured, with a mean percentage of silt and clay of 40.1 and 44.3%, respectively. Such variation of soil texture in both data sets is graphically illustrated in the USDA soil textural triangle (**Figure 2-4**) (e.g., the majority of soils samples falls within clay, silty clay, clay loam and silty clay loam classes; only few samples belong to other medium and coarse textured soil groups). An noticeable discrepancy of OC content between two data sets could probably be the outcomes of temporal and spatial variability among soil samples taken in these data sets.

In general, the soils in the study area are classified as Fluvisols, Gleysols, Luvisols, Acrisols, Arenosols and Plinthosols according to World Reference Base system (IUSS Working Group WRB, 2014), and correspond to Entisols, Inceptisols, and Ultisols of USDA/Soil Taxonomy system (Soil Survey Staff, 1975).

About soil structural evolution in VMD, it is widely recognized that soil structure development in such a young delta is strongly promoted by agricultural activities. As it was reported by Le (2003) and was observed during the soil campaign, the majority of soil samples taken from paddy fields showed massive structure in both surface plough layer and underneath compacted layers, while other soils cultivated upland crops showed some extent of structural development although in a weak grade. Since there was high OC accumulation on the plough layer of paddy rice, the hydro-physical properties of the massive horizons in paddy soil profile, however, are variation with depths, particulraly in the wet range of the SWRC due to the effect of organic carbon content on soil total porosity.

Table 2-2. Descriptive statistics of soil properties in the training (N=160) and test (N=29) data sets.

Soil properties	Training data set				Test data set			
	Min.	Max.	Mean	Std	Min.	Max.	Mean	Std
Organic carbon (%)	0.08	12.3	2.37	2.41	0.03	7.75	1.17	1.53
Bulk density (Mg m^{-3})	0.70	1.90	1.25	0.24	0.83	1.81	1.31	0.26
Sand content (%)	0.13	98.6	15.6	26.8	1.00	80.0	12.1	23.1
Silt content (%)	0.00	64.9	40.1	13.8	5.00	56.0	38.8	12.3
Clay content (%)	1.40	76.8	44.3	19.0	3.00	67.0	47.5	17.9
θ ($m^3 m^{-3}$) at:								
-1 kPa	0.24	0.74	0.50	0.10	0.31	0.66	0.52	0.08
-3 kPa	0.17	0.73	0.49	0.10	0.28	0.65	0.51	0.09
-6 kPa	0.12	0.72	0.47	0.12	0.26	0.62	0.50	0.09
-10 kPa	0.06	0.71	0.45	0.12	0.22	0.59	0.49	0.10
-20 kPa	0.03	0.70	0.41	0.12	0.17	0.55	0.46	0.10
-33 kPa	0.03	0.67	0.37	0.12	0.14	0.51	0.42	0.10

-100 kPa	0.03	0.58	0.32	0.11	0.08	0.43	0.35	0.09
-1500 kPa	0.02	0.43	0.24	0.09	0.04	0.25	0.21	0.06

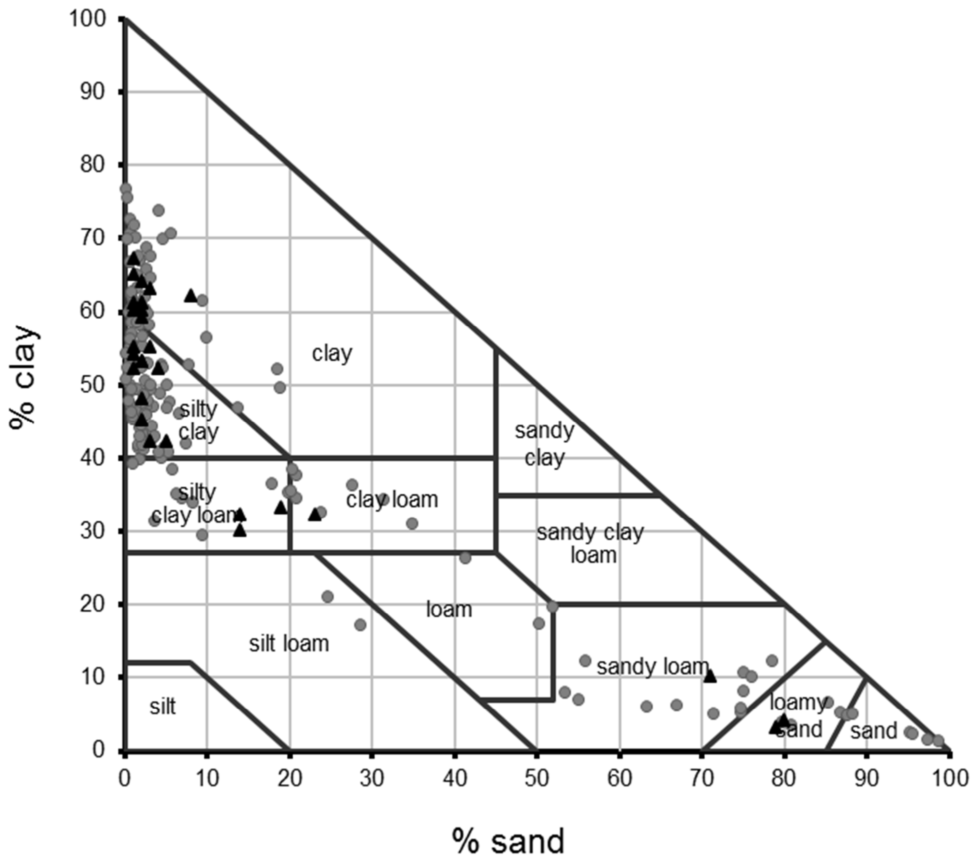

Figure 2-4. Variation of soil texture classes in the training data set (grey circles) and test data set (black triangles).

Chapter 3

PERFORMANCE EVALUATION OF PUBLISHED PTFS FOR PREDICTING SWRC OF VIETNAMESE MEKONG DELTA SOILS

This chapter is rewritten based on:

Phuong Minh Nguyen, Khoa Van Le, Yves-Dady Botula, Wim Cornelis (2015). Evaluation of soil water retention pedotransfer functions for Vietnamese Mekong Delta soils. *Agricultural Water Management* (158), 126-138.

3.1. Introduction

Soil water retention characteristic (SWRC), which expresses the functional relationship between soil matric potential and its corresponding gravimetric or volumetric water content (Or and Wraith, 2002), can be measured directly in the field or in the laboratory. The information of SWRC, however, is usually missing in most soil databases, due to its cumbersome and expensive measurement methods (Rawls and Brakensiek, 1982). The dearth of this information in developing countries located in the tropics is even worse than elsewhere due to additional problems linked to personnel training and acquisition of needed equipment for these measurements (Medina et al., 2002). In order to circumvent the missing gap, several efforts have been devoted to predicting SWRC from easily accessible soil properties using pedotransfer functions (PTFs).

Until recently, most of soil water retention PTFs available in the literature have been derived from soils in temperate regions (Tomasella and Hodnett, 2004), e.g., the PTFs of Gupta and Larson (1979); Lamorski et al. (2008); Nemes et al. (2008); Rawls and Brakensiek (1982); Saxton and Rawls (2006); Saxton et al. (1986); Schaap et al. (2001); Twarakavi et al. (2009); Vereecken et al. (1989). Much less studies have been devoted to soils in the tropics (Botula et al., 2012), particularly tropical humid deltas where paddy rice cultivation is the main agricultural practices. The lack of well-defined and extensive databases with hydraulic properties data is generally identified as the main constraint dragging the development of ‘tropical soils’ PTFs behind (Hodnett and Tomasella, 2002).

Since developing new PTFs is a very arduous task which generally requires a large soil database of good quality (Minasny et al., 2008a), using existing functions where possible is thus highly recommended. However, many PTFs have limited applicability, i.e., they have been derived for specific soils within a limited geomorphic and geographic domain and have been intended for a regional application. Specific PTFs might be accurate for the original training data, but unreliable for soils in other contexts (Wösten et al., 2001). As have been shown in several evaluation studies (e.g., Botula et al. (2012); Nebel et al. (2010)) the prediction performance of PTFs could be influenced by the geographic preference of the source data set. Additionally, Hodnett and Tomasella (2002) also cautiously noted the risks of applying PTFs developed using temperate soil databases to soils of the tropics. They observed marked differences between parameters which describe soil water retention behavior of soils in temperate and tropical climates. Such differences have been attributed to the discrepancy in terms of chemical, physical and mineralogical properties between soils. Indeed, although the soil forming factors are similar in both temperate and tropical climates, the extent of these factors is different (Lal, 2000). High prevailing temperature and intensive rainfall in the humid tropics result in strong weathering and leaching processes in large areas of the regions (loss of Ca, Mg, Na, K, and accumulation of Fe and Al) and tend to create particular minerals and soil structure that are less common in temperate regions (Hodnett and Tomasella, 2002). Therefore, Cornelis et al. (2001) and McBratney et al. (2002), among others, warned that the extrapolation of PTFs beyond the statistical limits of the calibration dataset and the geographical locations of soils from which they were developed should be avoided or at least carefully evaluated for their predictive quality.

Hence, it is important to test cautiously the applicability and predictability of published PTFs by using a limited number of measured SWRC from the site of interest (Espino et al., 1996; Minasny et al., 2008a). A major part of published evaluation studies in the literature have assessed PTFs developed for soils from temperate regions using independent data sets also from temperate climates (Buccigrossi et al., 2010; Cornelis et al., 2001; Kern, 1995; Tietje and Tapkenhinrichs, 1993). Only a limited number of studies have been conducted with the evaluated data sets of soils from humid and sub-humid tropics (Botula et al., 2012; Reichert et al., 2009; Tomasella and Hodnett, 2004). Additionally, Nemes et al. (2006a) noted that most evaluation studies of published PTFs using independent data sets in the literature remain unclear about the main sources of estimation errors. The difference between data sets used to derive PTFs, difference in the algorithms of PTF development, or difference among the predictors used might probably result in the overall error of prediction.

Therefore, the objectives of this study were to evaluate (1) the applicability and (2) reliability of a number of published SWRC-PTFs derived from soils in both temperate and tropical climates for the Vietnamese Mekong Delta (VMD) soils, and (3) to clarify the main sources of prediction error when using existing PTFs for soils in the tropical delta. To our knowledge, this is the first study focusing on evaluation of performance of existing PTFs for prediction of soil water retention property of a wide variety of soils in a delta dominated by paddy rice cultivation. This evaluation is important since it addresses the need of improving existing PTFs or developing new PTFs to offer more accurate estimations of SWRC in such regions.

3.2. Materials and methods

3.2.1. Evaluation data set

The evaluation was conducted in the Vietnamese Mekong Delta, whose detailed information about geo-pedo-hydro-climatic references was described in **Chapter 2**. The data set of 160 soil samples collected along the region was used for evaluating published PTFs' in this chapter. The distribution of soil properties in this data set can be referred in **Chapter 2**.

3.2.2. PTFs selected for evaluation

Some of the most commonly cited PTFs which were developed for soils in both temperate and tropical regions were selected for the evaluation, e.g., the 'temperate' PTFs of Gupta and Larson (1979); Rawls and Brakensiek (1982); Saxton and Rawls (2006), and the 'tropical' PTFs developed by Adhikary et al. (2008); Aina and Periaswamy (1985); Dijkerman (1988); Minasny and Hartemink (2011); van den Berg et al. (1997). Besides these widely cited PTFs, we also evaluated those of Salchow et al. (1996), since they were derived from alluvial soils similar to soils in our dataset, but under temperate climate, and that of Botula (2013) which was recently developed for soils in the humid tropic. All these PTFs, like a major part of PTFs available in the literature, are regression-based ones. However, there is an increasing interest nowadays in using data-mining or pattern-recognition techniques for PTF development. Therefore, in the present study, beside these well-known and recently developed regression-based PTFs, the widely cited PTFs of Schaap et al. (2001), based on Artificial Neural Networks (ANN) technique, and of Nemes et al. (2008), based on k-Nearest Neighbor (kNN) algorithm, were concomitantly selected for evaluation, using the user-friendly 'Rosetta' (Schaap et al., 2001) and 'k-Nearest' (Nemes et al., 2008) softwares, respectively.

The offered hierarchical PTFs in 'Rosetta' allow the estimation of van Genuchten water retention parameters due to the availability of limited (textural classes only) to more extended (texture, bulk density) input data. To evaluate the reliability of Rosetta for VMD soils, in this study, PTFs using sand, silt, clay percentage and bulk density as input variables were selected to get an estimation of SWRC parameters which were then used for calculation of soil water content at FC and PWP.

The ‘k-Nearest’ estimations of soil water retention at matric potentials of -33 kPa and -1500 kPa for VMD soils were obtained by using three different reference/training databases, such as (1) the default training database of the ‘k-Nearest’ software containing 2125 soils collected in the temperate climate from the NRCS-SCS Soil Characterization Database (Soil Survey Staff, 1997), (2) the IGBP-T data set of 534 soils from tropical regions withdrawn by Botula et al. (2013) from IGBP-DIS international database of ISRIC, and (3) a new data set of 196 soils from the Lower Congo used by Botula (2013) to derive corresponding regression-based PTFs. Sand, silt and clay content, bulk density and organic carbon content are input features used to determine the nearest neighbors of target soils.

The reasons of using three different training data sets to get the estimations of SWRC based on the kNN algorithm, in combination with other regression-based PTFs, are to examine the effect of (1) different soil databases derived from different climates and (2) different techniques used to develop PTFs based on the same data set on the overall prediction error for VMD soils.

All above-mentioned PTFs used typical soil properties that are routinely determined and available in soil survey databases (e.g., soil texture, bulk density and organic matter content) to predict SWRC at FC and PWP. Pedotransfer functions that needed more detailed information of input variables were not considered for evaluation in this study. Additionally, in the development of water retention PTFs, van den Berg et al. (1997) considered FC at -10 kPa matric potential. For this reason, the PTF of van den Berg predicting water content at -10 kPa was not considered in this evaluation study for FC.

The regression equations, the geographic domain, soil types and the range of soil properties from which selected PTFs were derived are summarized and displayed in **Table 3-1** and **Table 3-2**.

Table 3-1. List of selected regression-based PTFs developed for soils from both tropical and temperate regions and used to test for the VMD soils.

Source	PTFs	Soil types	Geographical domain
Tropical PTFs			
Aina and Periaswamy (1985)	$\theta_{-33\ kPa} = (0.6788 - 0.0055 \times Sa - 0.0013 \times BD \times Sa) \times BD$ $\theta_{-1500\ kPa} = (0.0213 + 0.0031 \times Cl) \times BD$	Alfisols, Ultisols.	Western Nigeria
Dijkerman (1988)	$\theta_{-33\ kPa} = (0.3697 - 0.0035 \times Sa) \times BD$ $\theta_{-1500\ kPa} = (0.0074 + 0.0039 \times Cl) \times BD$	Ultisols, Oxisols, Inceptisols	Sierra Leone
Adhikary et al. (2008)	$\theta_{-33\ kPa} = 0.5637 – 0.0051 \times Sa – 0.0027 \times Si$ $\theta_{-1500\ kPa} = 0.0071 + 0.0044 \times Cl$	various	India
van den Berg et al. (1997)	$\theta_{-1500\ kPa} = 0.00334 \times Cl \times BD + 0.00104 \times Si \times BD$	Oxisols and related soils	Global
Minasny and Hartemink (2011)	$\theta_{-33\ kPa} = 0.565 – 0.0749 \times BD - 0.0034 \times Sa$ $\theta_{-1500\ kPa} = 0.0795 + 0.0086 \times OC + 0.004 \times Cl - 0.00004 \times (Cl – 0.377)^2$	various	Tropical regions (ISRIC database)
Botula (2013)	$\theta_{-33\ kPa} = 0.4193 – 0.0035 \times Sa$ $\theta_{-1500\ kPa} = 0.0841 – 0.00159 \times Sa + 0.0021 \times Cl + 0.0779 \times BD$	Highly weathered soils	Lower Congo
Temperate PTFs			
Gupta and Larson (1979)	$\theta_{-33\ kPa} = 0.003075 \times Sa + 0.005886 \times Si + 0.008039 \times Cl + 0.002208 \times OM - 0.1434 \times BD$ $\theta_{-1500\ kPa} = -0.000059 \times Sa + 0.001142 \times Si + 0.005766 \times Cl + 0.002228 \times OM + 0.02671 \times BD$	Dredged sediments and soil materials	Eastern and central USA
Rawls and Brakensiek (1982)	$\theta_{-33\ kPa} = 0.2576 - 0.002 \times Sa + 0.0036 \times Cl + 0.0299 \times OM$ $\theta_{-1500\ kPa} = 0.026 + 0.005 \times Cl + 0.0158 \times OM$	various	USA
Salchow et al. (1996)	$\theta_{-33\ kPa} = -0.00064 \times Sa + 0.00123 \times Si + 0.00104 \times Cl + 0.02026 \times OM + 0.11421 \times BD$ $\theta_{-1500\ kPa} = -0.00126 \times Sa + 0.00039 \times Si - 0.00124 \times Cl + 0.03538 \times OM + 0.08426 \times BD$	Entisols, Inceptisols	Ohio, USA
Saxton and Rawls (2006)	$\theta_{-33\ kPa} = (\theta'_{-33\ kPa} + (1.283 \times (\theta'_{-33\ kPa})^2 – 0.374 \times \theta'_{-33\ kPa} - 0.015))/100$ $\theta'_{-33\ kPa} = -0.251 \times Sa + 0.195 \times Cl + 0.011 \times OM + 0.006 \times (Sa \times OM) - 0.027 \times (Cl \times OM) + 0.452 \times (Sa \times Cl) + 0.299$ $\theta_{-1500\ kPa} = (\theta'_{-1500\ kPa} + (0.14 \times \theta'_{-1500\ kPa} - 0.02))/100$ $\theta'_{-1500\ kPa} = -0.024 \times Sa + 0.487 \times Cl + 0.006 \times OM + 0.005 \times (Sa \times OM) - 0.013 \times (Cl \times OM) + 0.068 \times (Sa \times Cl) + 0.031$	various	Temperate regions (USDA database)

$\theta_{-33\ kPa}$ and $\theta_{-1500\ kPa}$: the volumetric water content ($m^3\ m^{-3}$) at field capacity (-33 kPa) and permanent wilting point (-1500 kPa); Sa, Si, Cl: sand, silt, clay content in decimal fractions ($kg\ kg^{-1}$); BD: bulk density ($Mg\ m^{-3}$), OM: organic matter content (%, OM = organic carbon (OC) content * 1.724)

Table 3-2. The range of soil attributes of the calibrated data sets used to derive the evaluated PTFs.

Source	No. of samples	Clay (%) (0-2µm)	Silt (%) (2-50µm)	Sand (%) (50-2000µm)	Organic matter (%)	Bulk density (Mg m^{-3})
Pattern recognition based-PTFs						
‘k-Nearest’ - NRCS (Nemes et al., 2008)	2125	0 - 81 (23)	0 - 92 (42)	0 - 81 (23)	1 - 14.9 (3.1)	0.5 - 1.9 (1.4)
‘k-Nearest’ - IGBP/T (Botula et al., 2013)	534	0 - 95 (36.5)	0.7 - 68.9 (22.8)	0.5 - 99 (40.1)	0 -13.7 (1.1)	0.4 - 1.9 (1.2)
Rosetta (Schaap et al., 2001)	2134	0 - 89	0 – 89	0 - 100	nm‡	0.4 – 2.0 (1.46)
Regression-based PTFs						
Aina and Periaswamy (1985)	48	8 - 43	6 - 20	42 - 86	nm	0.9 - 1.7
Dijkerman (1988)[a]	166	11.9 - 67.2	11.6 - 56.4	1.7 - 71.7	0.1 - 6.9	1.2 - 1.5
Adhikary et al. (2008)	1104	2.5 - 69.2	1 - 46	2.3 - 98.5	0.07 - 2	1.1 - 1.8
van den Berg et al. (1997)	91	10 - 95	nm	nm	0.17 - 9.0	0.8 - 1.6
Minasny and Hartemink (2011)	nm	0 - 90	nm	nm	nm	0.7 - 1.7
Botula (2013)	196	1.2 - 88.6 (44.6)	0.1 - 59.8 (15.6)	7.5 - 97.9 (43.7)	0.3 - 16.9 (3.9)	1.0 - 1.7 (1.4)
Gupta and Larson (1979)	nm	0 - 65	1 - 72	5 - 98	0 - 23	0.7 - 1.7
Rawls and Brakensiek (1982)	nm	1 - 93 (18)	1 - 93 (26)	1 - 99 (56)	0 - 7 (1)	0.7 - 2.1 (1.4)
Salchow et al. (1996)	108	9.8 - 34.4 (22.4)	23.5 - 66.4 (53.4)	4.4 - 64.7 (24.3)	1.3 - 4.4 (2.9)	1.1 - 1.8 (1.5)
Saxton and Rawls (2006)	1722	0 - 60	nm	0 - 99	0 - 8	1.0 - 1.8

[a]: information about the soil property range is found in Dijkerman and Miedema (1988) . Bulk density range was reported as average values of 1.2 Mg. m^{-3} for surface horizons (0-50 cm) and 1.5 for horizons deeper than 50 cm.

† Value between brackets is the arithmetic mean

‡ nm: information was not mentioned by the authors

3.2.3. Evaluation methodology

3.2.3.1. PTFs' applicability

Since major part of published PTFs, particularly the tropical PTFs, are site specific and have been developed for a limited range of soils; estimations of the output variables (i.e., SWRC at -33 kPa and -1500 kPa) were restricted to the range of independent variables (i.e., clay, silt, sand content, OC, BD) provided by the authors. Consequently, some PTFs were only applicable to a limited number of soil samples in evaluated data set. In the present study, the applicability of published PTFs to predict soil water retention at FC and PWP of VMD soils was evaluated using an applicability index. This index, expressed in percentage, is defined as the ratio between the number of time (i.e., number of soil samples in the evaluated data) that a particular PTF could be applied (in term of their inputs variables fall within the range of the PTF's calibration data set) and the total number of data points in the evaluated data set (Tomasella and Hodnett, 2004).

Although the values of some soil properties in our evaluated data set probably dropped outside the range of input variables provided by some PTFs, the evaluation of published PTFs in the present study was done on all soils of the test data set because our objective was to evaluate PTFs for a whole range of soils in the study area.

3.2.3.2. Evaluation criteria

Generally, the predictive capacity or reliability of different published PTFs for estimating SWRC of given soils can be evaluated based on the scatter plot of measured versus PTF predicted values of soil water content, in combination with several statistical indices. There are three complementary statistical indices which are often used to assess the performance of published PTFs: the mean error of prediction (ME – Eq. 3-1), which specifies an average tendency of overestimation or underestimation; the standard deviation of error of prediction (SDE – Eq. 3-2), which display the random variation of the predictions after correction of the bias, and the root mean square error of prediction (RMSE – Eq. 3-3) which denotes the overall error of prediction. These indices are formulated as:

$$ME = \frac{1}{N}\sum_{i=1}^{N}(E_i - O_i) \tag{3-1}$$

$$SDE = \sqrt{\frac{1}{N-1}\sum_{i=1}^{N}((E_i - O_i) - ME)^2} \tag{3-2}$$

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(E_i - O_i)^2} \quad (3\text{-}3)$$

where O_i is the i^{th} observed value in the evaluated dataset, E_i is the corresponding i^{th} estimated value obtained from selected PTFs, N is the number of observations in the evaluated data set.

3.3. Results and discussions

3.3.1. Exploratory data analysis

In order to get a thoughtful view about the interaction of soil properties in the test data set, preliminary data analysis was conducted to quantify the correlation strength between soil attributes, with particular attention to SWRC at FC (-33 kPa) and PWP (-1500 kPa) (**Table 3-3**).

Table 3-3. Pearson correlation matrix between soil variables in the evaluation data set (N=160).

	Sand	Silt	Clay	OC	BD	$\theta_{-33\ kPa}$	$\theta_{-1500\ kPa}$
Sand	1						
Silt	-0.74**	1					
Clay	-0.87**	0.33**	1				
OC	-0.28**	0.07	0.35**	1			
BD	0.45**	-0.23**	-0.47**	-0.75**	1		
$\theta_{-33\ kPa}$	-0.85**	0.53**	0.82**	0.49**	-0.61**	1	
$\theta_{-1500\ kPa}$	-0.87**	0.54**	0.84**	0.33**	-0.49**	0.92**	1

** Significant correlation at p<0.01

The Pearson correlation matrix displayed significant correlation between the volumetric soil water content at FC and PWP, and particle size distribution (sand, silt, clay content), dry bulk density (BD), and organic carbon (OC) content. Clay and OC content were positively correlated, while BD and sand content were negatively correlated to soil water content at FC and PWP. The correlation strength of BD and OC content with soil water content was higher at the field capacity than at the wilting point value, whilst the opposite was observed for clay and sand content. These observations are somehow logical because the structure of pore spaces which define the water content stored in soils at low suctions are more related to OC content and BD (Pachepsky et al., 2006), whereas soil water content at the nearly dry end of the SWRC is primarily determined by adsorption forces of the soil matrix (i.e. mainly determined by sand, silt, clay content and the mineralogy of clay fractions) (Manrique et al., 1991). Additionally, there is a significant negative relationship between soil OC content and BD (r = -0.75) confirming that soils with higher organic matter will concomitantly have lower bulk density.

3.3.2. Applicability of investigated PTFs

Concerning the overall applicability of the evaluated PTFs, **Table 3-4** presents the applicability index of each evaluated PTF for predicting SWRC at FC and PWP. Depending on which soil properties were selected as significant predictors, the applicability indices of PTFs for FC and PWP estimations might be different although they were derived by the same training database.

For the group of ‘tropical’ PTFs based on multivariate statistical regression technique, only those proposed by Minasny and Hartemink (2011) cover nearly the whole range of our evaluation dataset (98%), followed by the PTF of van den Berg et al. (1997) in which 82% of sampled soils in our evaluated dataset fit within the range of the calibration data set. The PTFs of Adhikary et al. (2008) which were derived from a national data set of India involving a wide range of major soil groups spread across the country, cover 35% and 55% of soils in our data set for FC and PWP estimation, respectively. Besides, the PTFs of Botula (2013) and of Dijkerman (1988), proposed for highly weathered soils in Lower Congo and Sierra Leone respectively, cover 28 and 22 %, as well as 23 and 33 %, respectively, of our soil samples. Since the PTFs of Aina and Periaswamy (1985) were developed for soils which were mainly formed from arenaceous sediments and contain predominantly kaolinite in the clay fraction, clayey soils and soils which contain high amounts of organic matter content in our evaluated data set were not covered by the calibration data sets of these PTFs. Hence, only 8 and 22 % of our evaluated data set falls within the range of the calibration data set of these PTFs.

In the groups of ‘temperate’ PTFs, the PTFs of Rawls and Brakensiek (1982), and Saxton and Rawls (2006) cover respectively 85 and 72 % of our soil samples. The PTFs of Gupta and Larson (1979) excluded soils with very low sand content (< 5 %), or relatively high clay content (> 65 %), thus clayey soils that have clay content larger than 65% in our data set fall out of the range of their calibration data set. In case of the PTFs of Salchow et al. (1996), only 3 % of our data set was covered by their calibration data set. The very low applicability index of Salchow et al. (1996) PTFs resulted from the fact that the soils they used to derive the regression equations are alluvial soils of coarse and medium textures (i.e. silty clay loam, silt loam, loam and sandy loam). Thus, predominant fine-textured soils in our data set did not match the ranges of these soils.

In case of pattern recognition-based PTFs, the Rosetta PTFs of Schaap et al. (2001), which were calibrated by assembled large databases of soils in temperate and subtropical climates of North America and Europe, completely cover the whole range of soils in the evaluated data set in term of input attributes. Similarly, kNN-PTFs using the general and large reference/training databases of soils from both temperate (NRCS database, Nemes et al. (2008)) and tropical climates (IGBP/T database, Botula et al. (2013)) are highly applicable to VMD soils with an applicability index of 83 % and 87 %, respectively. On the contrary, the kNN-PTFs using a local data set of Lower Congo soils (Botula, 2013) cover around 20% of the evaluated data set. Since this kNN estimation is based on the inputs of particle size distribution, BD, and OC; its applicability index is different from the regression-based PTFs derived by the same data set.

Briefly, the PTFs derived from international databases were highly applicable to Mekong Delta soils, at least in coverage of input data. Lilly and Lin (2004) also claimed that PTFs should be used as interpolation tool to predict a desired soil property belonging to the range from whose data the PTF was developed. However, Tranter et al. (2009) pointed out that although ‘international database’ PTFs offered greater coverage, they are far less precise than those trained on less diverse data. Moreover, since the applicability index was calculated based on solely input features which were actually used in particular PTFs (i.e., sand, silt, clay content, bulk density and organic matter content), other soil properties (e.g., soil structure and clay mineralogy) which are expected to play a major role in SWRC, are not reflected by this index. The index, therefore, cannot assure the good performance of published PTFs on the soils of interest. This hypothesis will be tested in the section below in term of other evaluation indices (e.g. ME, SDE, RMSE).

3.3.3. Reliability of evaluated PTFs

The indices ME, SDE and RMSE which are considered as measures for the predictive quality, or ‘reliability’, of existing PTFs are presented in **Table 3-4**. Concurrently, the performance of evaluated PTFs in predicting volumetric water content at FC and PWP of soils in the VMD is graphically depicted in **Figure 3-1**, **Figure 3-2**, and **Figure 3-3**.

It can be seen that the investigated PTFs in this study had contrasting predictive performance for VMD soils. The results of RMSE, which was used as the main criterion for evaluation, showed that for predicting soil water retention at FC, the 'tropical' regression-based PTFs of Adhikary et al. (2008), and Botula (2013) together with 'k-Nearest' PTFs based on the 'tropical' IGBP-T database and the 'temperate' NRCS/SCS database performed best (with RMSE values in the range of 0.06 – 0.064 m^3 m^{-3}), followed by the 'tropical' PTF of Minasny and Hartemink (2011) (RMSE = 0.072 m^3 m^{-3}), and the 'temperate' PTF of Saxton and Rawls (2006) (RMSE = 0.073 m^3 m^{-3}). For estimating soil water retained at PWP, all aforementioned PTFs provided consistent performance with RMSE ranging from 0.049 to 0.063 m^3 m^{-3}. On the contrary, the PTFs of Aina and Periaswamy (1985), Dijkerman (1988), Gupta and Larson (1979), Rawls and Brakensiek (1982), Salchow et al. (1996), were not successful in predicting SWRC at both FC and PWP for VMD soils (RMSE in the range of 0.103 – 0.370 m^3 m^{-3} for FC and 0.070 – 0.140 m^3 $m^{-3}$for PWP). Although the pattern recognition PTFs of 'k-Nearest' using Lower Congo data and 'Rosetta' PTFs performed adequately at PWP (RMSE = 0.060 m^3 m^{-3}), they did not provide accurate estimation of SWRC at FC (RMSE = 0.081 - 0.090 m^3 m^{-3}).

Further detailed assessment about the contribution of erroneous components, i.e. accuracy error - ME, and precision error – SDE, on the total prediction error of investigated PTFs revealed that most of 'tropical' PTFs derived based on the soils of the tropics, except one, display a little bias in predicting water retention of VMD soils with absolute values of ME ranging from 0.002 m^3 m^{-3} to 0.046 m^3 m^{-3} for -33 kPa matric potential and from 0.007 m^3 m^{-3} to 0.039 m^3 m^{-3} for -1500 kPa matric potential. The exception was the regression-based PTFs of Aina and Periaswamy (1985) which substantially overestimates soil water content at FC (ME = 0.31 m^3 m^{-3}) and relatively underestimate it at PWP (ME = -0.05 m^3 m^{-3}). Such huge bias could probably result from their small training dataset (N = 48 observations), the difference in methods used to determine soil texture by these authors (hydrometer method) as compared to the present study (sieve and pipette method), as well as the difference of soil origin between their calibration data set (arenaceous sediments) and our test data set (fluvial sediments).

Table 3-4. Applicability index and statistical indices (ME, SDE, RMSE) yielded by the investigated PTFs.

Evaluated PTFs	Applicability index (%)	$\theta_{-33\ kPa}$			$\theta_{-1500\ kPa}$		
		ME	SDE	RMSE	ME	SDE	RMSE
Regression based PTFs							
Tropical climates							
Aina and Periaswamy (1985)	8 – 23	0.307	0.209	0.371	-0.050	0.064	0.081
Dijkerman (1988)	23 – 33	0.009	0.116	0.116	-0.024	0.066	0.071
van den Berg et al. (1997)	82	-	-	-	-0.012	0.065	0.066
Adhikary et al. (2008)	35 – 56	0.002	0.060	0.060	-0.039	0.049	0.063
Minasny and Hartemink (2011)	98	0.044	0.057	0.072	0.020	0.046	0.050
Botula (2013)	28 – 22	-0.009	0.063	0.064	0.007	0.051	0.049
Temperate climates							
Gupta and Larson (1979)	31	0.096	0.059	0.113	0.102	0.060	0.118
Rawls and Brakensiek (1982)	85	0.134	0.119	0.179	0.071	0.087	0.112
Salchow et al. (1996)	3	-0.063	0.081	0.103	-0.050	0.131	0.140
Saxton and Rawls (2006)	72	-0.006	0.073	0.073	0.017	0.053	0.056
Pattern-recognition based PTFs							
Rosetta (Schaap et al., 2001)	100	-0.060	0.055	0.081	-0.025	0.065	0.060
k-Nearest /NRCS (Nemes et al. , 2008)	83	-0.014	0.062	0.063	0.030	0.055	0.063
k-Nearest/IGBP-T (Botula et al., 2013)	87	-0.012	0.063	0.064	0.022	0.048	0.052
k-Nearest/Lower Congo (Botula, 2013)	20	-0.046	0.091	0.090	-0.008	0.060	0.060

$\theta_{-33\ kPa}$ and $\theta_{-1500\ kPa}$: the volumetric water content ($m^3\ m^{-3}$) at field capacity (-33 kPa) and permanent wilting point (-1500 kPa). ME is the mean error of prediction ($m^3\ m^{-3}$), SDE is the standard deviation of error of prediction ($m^3\ m^{-3}$), RMSE is the root mean square error of prediction ($m^3\ m^{-3}$).

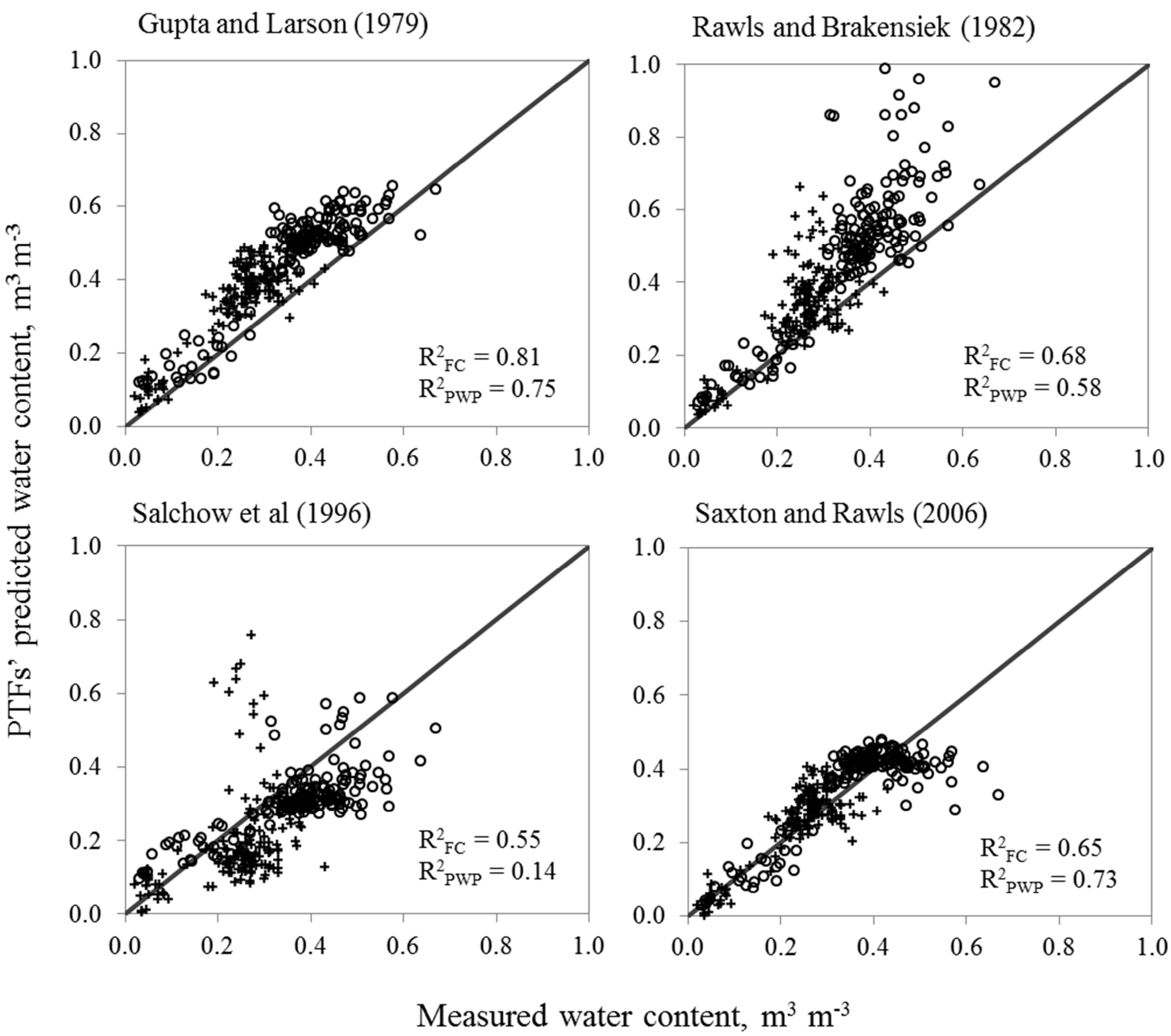

Figure 3-1. Correspondence of measured vs. PTFs' predicted soil moisture content ($m^3\ m^{-3}$) at FC, -33 kPa (circles) and at PWP, -1500 kPa (crosses) of regression based PTFs developed for soils in temperate climates.

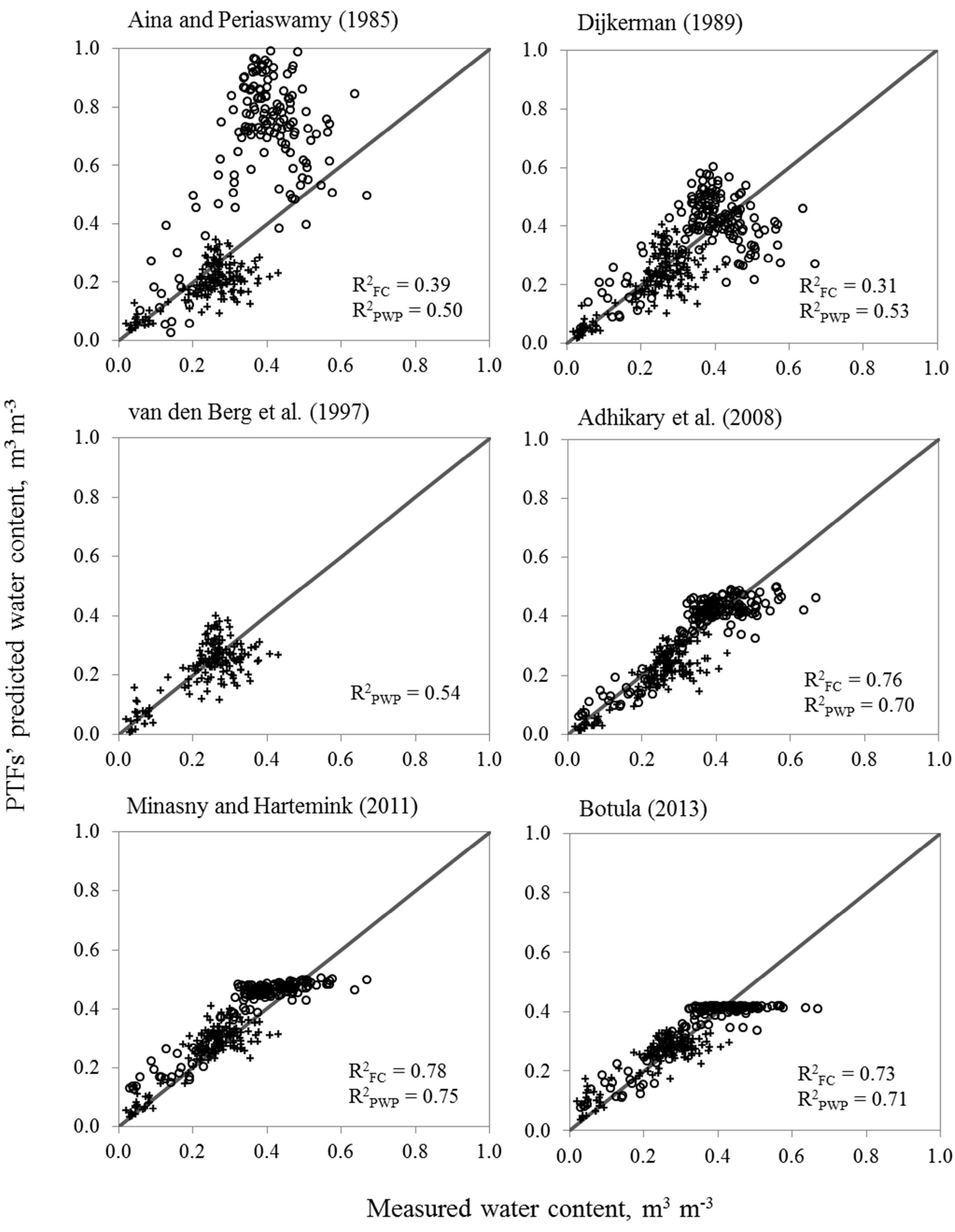

Figure 3-2. Correspondence of measured *vs.* PTFs' predicted soil moisture content (m^3 m^{-3}) at FC, -33 kPa (circles) and at PWP, -1500 kPa (crosses) of regression based PTFs developed for soils in tropical climates.

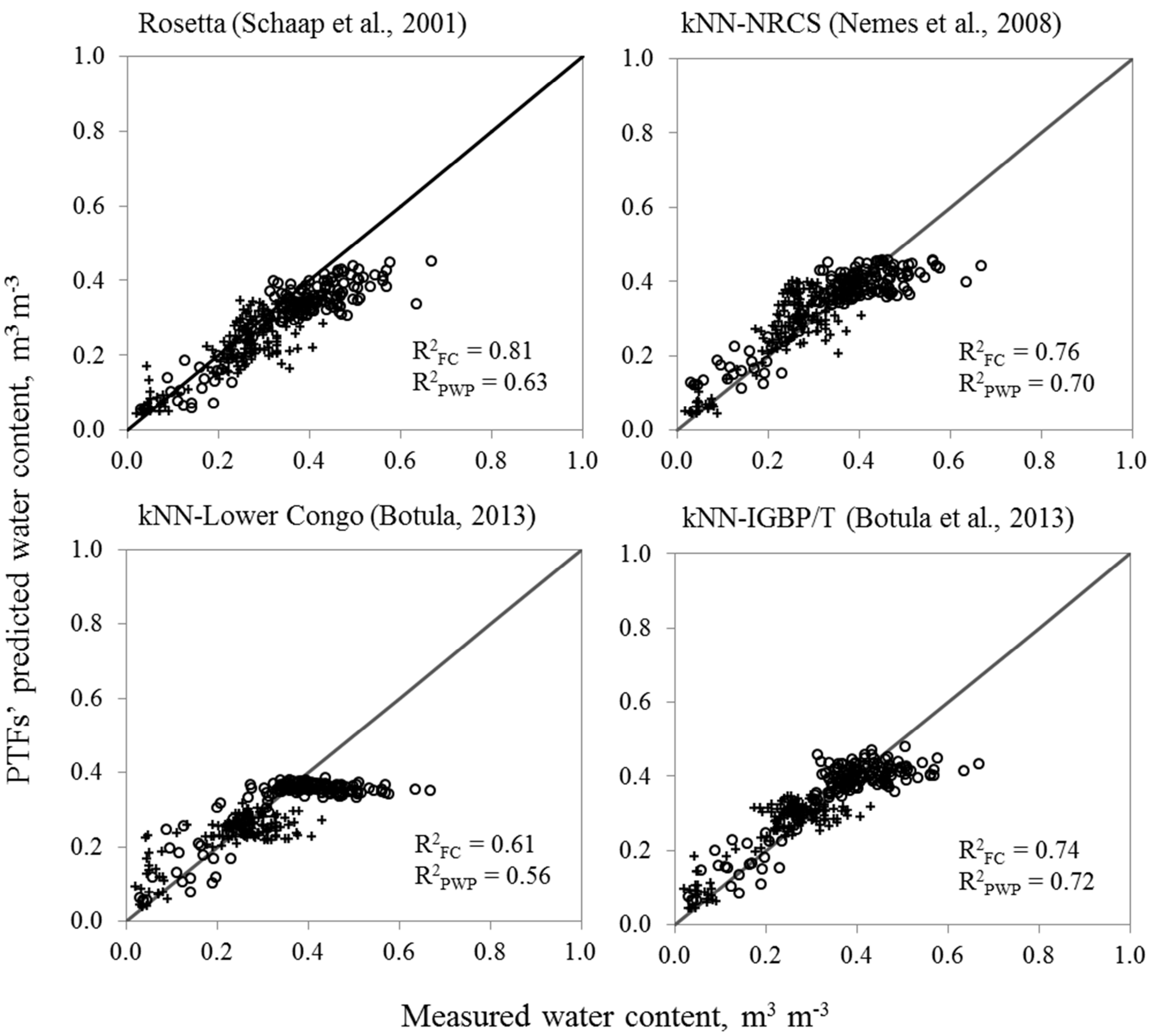

Figure 3-3. Correspondence of measured *vs.* PTFs' predicted soil moisture content (m^3 m^{-3}) at FC, -33 kPa (circles) and at PWP, -1500 kPa (crosses) of pattern-recognition based PTFs.

Consequently, a large proportion of the total prediction error of the 'tropical' PTFs was explained by the precision error (SDE). Wide scattering of the measured versus PTF predicted points around the 1:1 reference line (**Figure 3-2** and **Figure 3-3**), corresponding to high SDE values, suggest the presence of other factors affecting soil water retention at FC and PWP of VMD soils that might probably be not accounted for in the evaluated PTFs. These results are no surprising, because most of investigated 'tropical' PTFs in the present research were derived from highly weathered soils which are dominated by low activity clays (e.g. kaolinite and sesquioxides) and low organic matter content. Indeed, the differences in the water retention of different soil minerals (e.g., kaolinite, illite and montmorillonite) are proved to be more significant at FC (or low suctions) than at PWP (Tessier et al., 1992). This explains why the 'k-Nearest'-PTFs using Lower Congo soils performed well in predicting water retention at PWP, but not at FC.

On the other hand, the regression-based PTFs derived based on soils from temperate regions of Gupta and Larson (1979), Rawls and Brakensiek (1982), and Salchow et al. (1996) produced less accurate (with absolute values of ME in the range of 0.063 - 0.134 m^3 m^{-3} at -33 kPa and of 0.05 - 0.102 m^3 m^{-3} at -1500 kPa) and imprecise prediction (SDE from 0.059 to 0.119 m^3 m^{-3} at FC and from 0.06 to 0.131 m^3 m^{-3} at PWP) for soils in the tropical delta. The sign of ME (**Table 3-4**) together with the scatter plot (**Figure 3-1**) shows that PTFs of Gupta and Larson (1979), and Rawls and Brakensiek (1982) overestimate the soil water retained at both FC and PWP, whilst the PTFs of Salchow et al. (1996) tend to underestimate water content at -33 kPa and -1500 kPa (**Table 3-4**). Undoubtedly, different soil texture classes have completely different soil water retention characteristics. 'Temperate' PTFs empirically derived from databases with bias to coarse- and medium-textured soils (e.g., PTFs of Gupta and Larson (1979), Rawls and Brakensiek (1982), and Salchow et al. (1996)) offered unsatisfactory estimation of water retention, especially at PWP, for predominantly fine-textured soils in our data set. The only noticeable exception of 'temperate' PTFs was the ones of Saxton and Rawls (2006). These PTFs showed modest prediction quality at both FC (RMSE = 0.073 m^3 m^{-3}) and PWP (RMSE = 0.056 m^3 m^{-3}). A small caution note when using the equations of Saxton and Rawls (2006) for estimating soil water retention characteristics is the unit of predictor variables in the regression equations. In their paper, they mentioned sand, silt, clay content in weight percentage, but actually one has to use the decimal fraction of particle size distribution to get the right estimation.

In general, many of 'tropical' PTFs (e.g., PTFs of Adhikary et al. (2008), Minasny and Hartemink (2011), Botula (2013), and kNN-PTFs using tropical IGBP database) are far more reliable than 'temperate' PTFs in estimating soil water retention at FC and PWP of soils in VMD, although the limited predictive potentials of these PTFs in describing the SWRC of young alluvial soils in VMD were also recognized in the present evaluation. Since the shape of the SWRC in the portion of high matric potentials is mainly defined by soil structure and clay mineralogy (Bruand, 2004b; Kay and Angers, 2002); 'tropical' PTFs biasedly derived from highly weathered soils (e.g. soils with strong micro-aggregated structure and dominated by sesquioxides and kaolinite), have some restraints in describing accurately the SWRC, particularly in the wet and intermediate range, of soils in VMD.

As large proportions of soils in temperate regions are also relatively young (Minasny and Hartemink, 2011), the difference in clay mineralogy might not be the main cause of inadequate performance of 'temperate' PTFs on VMD soils. The variation in the distribution of soil properties, particularly soil texture, between the PTFs' calibration databases and test data set seems much more pronounced. The soils in temperate regions are much biased to sandy and loamy textural classes (Hodnett and Tomasella, 2002), while the tropical delta soils in this study have a far higher proportion of clayey textural classes. It is basically acknowledged that the whole SWRC is influenced by soil texture, and the 'temperate' PTFs therefore have a limited capability to predict SWRC, particularly at PWP, of clayey soils in VMD.

3.3.4. Sources of prediction errors

3.3.4.1. Quality of PTFs' calibration data sets

The results of the evaluation study presented in the above sections highlighted the strong influence of PTFs' calibrated data sets, in terms of their size, their coverage (expressed by the applicability index) and their similarity in soil-forming factors such as climate (defined by geographical position i.e., tropical or temperate regions), time (e.g., highly weathered soils *vs.* alluvial soils with recent origin), and soil textural class (sandy, loamy, clayey soils), to the predictive performance of published PTFs. Obviously, investigated PTFs derived from large soil databases which cover a wide range of soils in the tropics (e.g., PTFs of Adhikary et al. (2008), Minasny and Hartemink (2011), kNN-PTFs using tropical IGBP database) provide more adequate prediction of SWRC at FC and PWP than the others. The result of the correlation test showed that the applicability index and the error measure (i.e. RMSE) in this evaluation study are significantly correlated, with a Kendall's Tau-b value of -0.28 ($p<0.05$). There is a tendency that increasing applicability index results in a decreasing RMSE. These results are in good accordance with those reported by Botula et al. (2012); Medeiros et al. (2014); Nebel et al. (2010). These authors contended that the performance of PTFs could be influenced by the size and the origin of the PTF's training data sets.

3.3.4.2. Development methods

Having compared the predictive quality of published PTFs for tropical delta soils, it is reasonable to look at the effect of regression methods used to derive PTFs on their predictive quality. It has been shown that novel data-mining methods generate more flexible and accurate PTFs than statistical regression (Botula et al., 2013; Haghverdi et al., 2012; Skalová et al., 2011). Our research partially agrees with these findings with adequate performance of 'k-Nearest'-PTFs using international databases from tropical (IGBP-T) and temperate (NRCS) climates. However, it is important not to overemphasize the supremacy of data mining techniques in all cases, since the better performance of novel data-mining methods strongly depend on the quality and the homogeneity of data set used to train and test the model (Haghverdi et al., 2014). Evidence can be observed when comparing the performance of regression-based PTFs of Botula (2013) and kNN-based PTFs using the Lower Congo data set as reference. These two PTFs use the same training data set to derive the predictive models and regression-based PTFs outperformed the 'k-Nearest'-PTFs. The worse performance of the latter (RMSE = 0.09 $m^3 m^{-3}$ at FC and 0.06 $m^3 m^{-3}$ at PWP) was strongly explained by the sensitivity of kNN, a similarity-based technique, to the difference between soils in the reference and test data sets, since the prediction of test samples was obtained based on the observations in the reference data set.

3.3.4.3. Used predictors

Another source of overall prediction error when implementing the existing PTFs is the difference in predictors used to obtain the estimation. In this study, we clarify that using more predictors does not necessarily assure a successful performance of existing PTFs for a given soil. As we have observed, most of the investigated ‘tropical’ PTFs use only particle size distribution and BD, and yet they generally offer quite accurate estimations of SWRC for VMD soils compared to the investigated ‘temperate’ PTFs which exploited particle size distribution (sand, silt, clay content), OC content and BD as predictor variables. Preliminary data analysis about the correlation strength of soil attributes in the evaluation data set displayed a strong correlation between the soil water content at FC and PWP, and soil texture, BD and OC (**Table 3-3**). Moreover, because of the strong correlation between BD and OC, using soil BD as predictor indeed would provide virtually the same information as OC does in characterizing SWRC of VMD soils.

In brief, the characteristics of the PTFs’ calibration data sets might probably be the leading cause of prediction error when applying published PTFs to VMD soils. In order to select the suitable PTFs from the literature for soils in a specific region of interest, we suggest to use the applicability index together with geographic position as integral indicator for PTFs selection. This suggestion, however, is based on the observations of this study, and it remains to be seen whether the proposed integral indicator is valid and holds for other cases.

3.4. Conclusions

This study shows that the predictive performance of published PTFs for tropical delta soils varies and depends on the coverage and the quality of PTF’s calibration data sets. In order to adequately describe SWRC of tropical VMD soils, the calibration data set of published PTFs should be large and even representative to account for variability of soil properties in the region of interest. Several features e.g., applicability index and geographic conditions are expected to provide preliminary information related to the suitability of particular PTFs for soils at specific regions. They could therefore be used as integral indicator to select appropriate PTFs in cases no data of SWRC and no specific PTFs are available for timely uses.

Moreover, detailed evaluation of PTFs' performance additionally revealed that PTFs developed based on 'tropical' data sets generally perform better than those using 'temperate' data sets for predicting water content at FC and PWP for VMD soils. However, their predictive potential for VMD soils remains limited due to their specific soil genesis together with representative management of paddy rice cultivation in such regions. Further research is needed to develop specific PTFs for tropical delta soils using extensive databases collected from tropical delta regions. Using these specific PTFs could provide more accurate estimations of SWRC for subsequently uses in simulation models related to agricultural and environmental studies (e.g., AquaCrop, HYDRUS, and many others).

Chapter 4

THE USE OF CATEGORICAL SOIL STRUCTURE INFORMATION IN DEVELOPING SOIL WATER RETENTION PTFS

This chapter is written based on:

Phuong Minh Nguyen, Khoa Van Le, Wim M. Cornelis (2014). Using categorical soil structure information to improve soil water retention estimates of tropical delta soils. *Soil Research* 52(5): 443-452.

4.1. Introduction

Except for the availability of large databases, which boosted the development of PTFs in temperate regions, there are few well-documented and exhaustive databases for soils in the tropics (Minasny and Hartemink, 2011). That limitation in the tropical regions constrains the development of hydraulic PTFs and leads to difficulties in applying water and solute transport simulation models in those regions.

The value of soil structure in predicting soil hydraulic properties has been widely documented in literature (Coen and Wang, 1989; Jong and McKeague, 1987; King and Franzmeier, 1981; Lilly, 2000; Lilly et al., 2008; McKeague et al., 1982; McKenzie and Jacquier, 1997; Pachepsky and Rawls, 2003; Williams et al., 1992). Soil structure provides information about pore architecture and, consequently, the soil pore volume in which the water is held. Therefore it would seem logical to use the visual field description of soil structure to predict soil water retention (Lilly and Lin, 2004). Soil structure information is typically available in many soil survey databases, and has been reported as a potential predictor or promising grouping criterion in developing SWRC-PTFs (Abbaspour and Moon, 1992; Bruand, 2004a; Danalatos et al., 1994; McKenzie and MacLeod, 1989; Rawls and Pachepsky, 2002; Williams et al., 1992). However, despite the substantial volume of soil structural data existing in soil survey databases, such soil information still remains under-utilized in PTFs' development (Calhoun et al., 2001; Lilly and Lin, 2004).

Grouping can be done with different criteria such as genetic horizons, texture, bulk density, soil structure, parent materials, and others (Bruand, 2004a; Danalatos et al., 1994; Pachepsky and Rawls, 1999; Williams et al., 1992; Wösten et al., 1995; Wösten et al., 1999). Most of the studies found that grouping improved the accuracy of predictive functions, but none of the above properties could be considered as the best grouping criteria. Selecting properties as grouping criteria should account for specific agro-morphological characteristics of the soils in the study area.

During the past decade, considerable progress has been made in developing hydraulic PTFs for tropical soils, as is illustrated by studies of Adhikary et al. (2008); Aimrun and Amin (2009); Botula et al. (2013); Mdemu and Mulengera (2002); Minasny and Hartemink (2011); Obalum and Obi (2012); Patil et al. (2013); Shwetha and Varija (2013); Suprayogo et al. (2003). However, few efforts were devoted to tropical delta soils.

In the tropical Mekong Delta of Vietnam, where this study was conducted and where the main agricultural practice is paddy rice cultivation, the soil is usually prepared under submerged conditions generating the typical massive plough layer. The physical and hydraulic soil characteristics of the puddled layers are tremendously different when compared to those under other land uses, especially upland crop cultivation (Kögel-Knabner et al., 2010; Linh et al., 2014).

Because of the very specific nature of soil properties and representative agricultural practices in tropical deltas, it is not advisable to estimate the soil water characteristics in these regions by utilizing PTFs reported so far in the literature (as manifested in **Chapter 3**). The objective of this study was therefore to develop predictive functions to estimate the soil water retention characteristics (SWRC) of Vietnamese Mekong Delta (VMD) soils based on basic soil properties that are usually available in many soil data sets (e.g., soil texture, bulk density and organic matter content), and to investigate whether including soil structure on top of these widely used predictor variables can improve the accuracy of prediction. Additionally, since SWRC-PTFs in the tropics are now in the development stage, we examined the improved effect of some supplementary soil properties available in our data set, such as pH, electrical conductivity (EC), soil plastic limit, and stability index on our PTFs predictability. Soil plastic limit has been reported as a promising predictor of SWRC estimation due to the strong correlation between plastic limit and SWRC as well as with others soil properties (e.g., soil texture and organic carbon content) (Dexter and Bird, 2001; Khlosi et al., 2013). To our knowledge, this is the first study focusing on prediction of soil water retention characteristics of a wide variety of soils in a delta dominated by paddy rice cultivation, and in which specific attention is given to consideration of soil structure information.

4.2. Materials and Methods

4.2.1. Soil data set

The data set of 160 samples was utilized to investigate the predictive power of soil properties for SWRC estimations. Descriptive statistics of soil variables in this data set were presented in **Chapter 2**.

4.2.2. Predicting soil water retention

Multiple linear regression (MLR) analysis, performed by the statistical package SPSS v. 20 (IBM, 2011), was employed to develop predictive equations for predicting the SWRC. It is because MLR enables to automatically detect essential input variables using a stepwise regression method (Rawls et al., 1991). Moreover, Wösten et al. (2001) illustrated that better and larger number of soil data are expected to improve the performance and accuracy of predictive functions rather than exploiting complicated techniques. Recently, Minasny and Hartemink (2011) also used the MLR technique for developing PTFs for predicting BD and water content at FC and PWP of tropical soils.

Point PTFs enabling to predict water content at specific pressure heads (eight heads in this study) were developed based on the data set of 160 observations with classical independent variables (soil texture, OC content and soil BD). After that, the above data set were partitioned by “grouping” (Bruand, 2004a), into three homogeneous subsets of structured, massive and structureless soils (N = 46, N = 91 and N = 23 observations, respectively). The partition was done by using the information of soil structure description from the soil survey. According to FAO guidelines for visual soil structure assessment (FAO, 2006), the soil was first examined and divided into apedal soil (lacking of soil structure) and pedal soil (showing soil structure, and thus named “structured” soils in the present study). Apedal soils were then subdivided into single grain (i.e., “structureless” soils) and “massive” soils. The three recorded specific-structure soil groups were used to test whether incorporating categorical soil structural information could improve the prediction accuracy of regression equations. Additionally, the explanatory effect of supplementary variables on the variability of SWRC was evaluated by using an extended set of predictors (i.e. one including classical predictors plus other sometimes available or easily measurable soil properties as pH, EC, soil plastic limit and aggregate stability index) to build up PTFs.

4.2.3. Evaluating the accuracy of prediction

There are several indices used in the literature to evaluate the accuracy of indirect methods, such as the coefficient of determination (R^2), the mean error (ME), the absolute mean error (AME), the root mean square error (RMSE), and the unbiased root mean square error (URMSE) (Schaap, 2004). Among them, R^2, ME and RMSE are the most commonly used. In this study, since the regression coefficients of developed PTFs were calibrated using the least-square method as optimization function, the ME is almost close to zero. Therefore, to evaluate the accuracy of our newly developed PTFs, we used the RMSE and the R^2 indices. The RMSE is the measure of the overall prediction error and R^2 refers to the strength of the linear relationship between measurement and prediction, which indicates the amount of variability explained by the regression equation. The equations for R^2 and RMSE calculation were presented in **Chapter 3.**

4.3. Results and Discussions

4.3.1. Performance of point PTFs developed based on the whole dataset

Preliminary data analysis displayed the exponential relationship between total organic carbon content and soil moisture content retained at different pressure heads. We therefore applied a log-transformation to resolve the non-linearity problem of total organic carbon content before conducting multiple linear regression (MLR) analysis. The correlation strength between our available independent variables and soil water retention at eight pressure heads is illustrated in **Table 4-1**.

The Pearson's coefficients in **Table 4-1** reveal the strong relationship between soil moisture retained at different matric potentials and five promising predictors, i.e. logarithmic organic carbon content (log(OC)), plastic limit (PL), bulk density (BD), clay (Cl) and sand (Sa) content. Log(OC), PL and Cl were positively correlated with soil water content, while BD and Sa were negatively correlated. The correlation strength between log(OC), PL and BD and soil moisture decreased with increasing soil matric suctions, whilst those of clay and sand content rose with increasing suctions. These observations are logically desirable, because the structure of pore spaces which define the water content stored in soils at low suctions is more related to organic carbon content and bulk density (Pachepsky et al., 2006), whereas soil water content at the dry end is primarily determined by adsorption forces in the soil matrix (mainly determined by the quantity of clay, silt, and sand particles) (Manrique et al., 1991). Botula et al. (2012) also observed similar correlation trends of water content at FC and PWP with sand, clay content and BD of tropical Lower Congo soils.

Table 4-1. Pearson's correlation between potential predictors and measured soil volumetric water retention.

Water content	pH	EC	Log(OC)	BD	PL	Sa	Si	Cl	SI
$\theta_{-1\ kPa}$	-0.449**	0.237**	0.753**	-0.722**	0.780**	-0.638**	0.268**	0.708**	0.152
$\theta_{-3\ kPa}$	-0.426**	0.254**	0.772**	-0.717**	0.803**	-0.683**	0.306**	0.743**	0.11
$\theta_{-6\ kPa}$	-0.487**	0.273**	0.799**	-0.677**	0.828**	-0.771**	0.406**	0.795**	0.01
$\theta_{-10\ kPa}$	-0.499**	0.283**	0.796**	-0.653**	0.830**	-0.815**	0.416**	0.818**	-0.022
$\theta_{-20\ kPa}$	-0.490**	0.293**	0.756**	-0.630**	0.818**	-0.852**	0.508**	0.835**	-0.039
$\theta_{-33\ kPa}$	-0.469**	0.291**	0.713**	-0.605**	0.786**	-0.855**	0.531**	0.823**	-0.041
$\theta_{-100\ kPa}$	-0.443**	0.282**	0.65**	-0.562**	0.727**	-0.854**	0.559**	0.801**	-0.044
$\theta_{-1500\ kPa}$	-0.417**	0.255**	0.598**	-0.493**	0.713**	-0.870**	0.536**	0.84**	-0.044

** Significant correlation at 0.01 significance level.

pH is pH_{H2O} measured in 1:2.5 ratio of dilution, EC (mS cm^{-1}) is the electrical conductivity measured in 1:2.5 ratio of dilution, log(OC) is the logarithmic form of total organic carbon content (%), BD is soil bulk density (Mg m^{-3}), PL is plastic limit (kg kg^{-1}), Sa is sand content (%), Si is silt content (%), Cl is clay content (%), SI is stability index, θ is volumetric water content (m^3 m^{-3}) at different matric potentials.

Based on the above correlation relationships between SWRC and independent soil variables, eight point-PTFs were developed by stepwise regression technique (with probability $p < 0,05$ for entry and $p < 0.01$ for removal of predictor). The result shows that SWRC of tropical Mekong delta soils can be satisfactorily estimated by soil texture, total organic carbon content and bulk density. The selected predictors and their corresponding regression coefficients of point PTFs are presented in **Table 4-2**.

Table 4-2. Regression coefficients and coefficient of determination of point-PTFs.

Matric potential (kPa)	Regression coefficient [a]						R^2
	Intercept (10^{-2})	Sand (10^{-3})	Silt (10^{-3})	Clay (10^{-3})	Log(OC) (10^{-2})	BD (10^{-2})	
-1	57.5			2	5.5	- 14.4	0.71
-3	52.7			2	6.7	- 12.5	0.75
-6	36.7		1	3	12	- 6.2	0.80
-10	22.8		1	3	12.7		0.83
-20	41.5	- 2		2	6.6	- 5.8	0.89
-33	49.3	- 2		1		- 11.8	0.83
-100	49.7	- 3				- 10.7	0.80
-1500	23.4	- 2		2		- 3.2	0.82

[a] Sand, Silt, Clay, and OC (organic carbon) content are in weight percentage; BD (bulk density) is in Mg m^{-3}.

The coefficient of determination (R^2) of our point-PTFs showed that most of the variability of SWRC (ranging from 71 % to approximately 89 %) can be explained by regression equations which exploited classical soil properties as predictors. Moreover, the graphical representation of measured versus PTF predicted soil water content with the 1:1 reference line (**Figure 4-1**) also strengthens the appropriateness of these point PTFs in describing the SWRC. There is a good agreement between practical measured and PTF predicted values as most of the points are closely scattered around the reference line, and do not exhibit much bias. Our findings confirm the results of Tomasella et al. (2003); Vereecken et al. (2010); Wösten et al. (2001). These authors point out that moisture content at different pressure heads is controlled by different soil properties and, therefore, point PTFs should provide a better combination of these properties leading to more accurate functions for estimation.

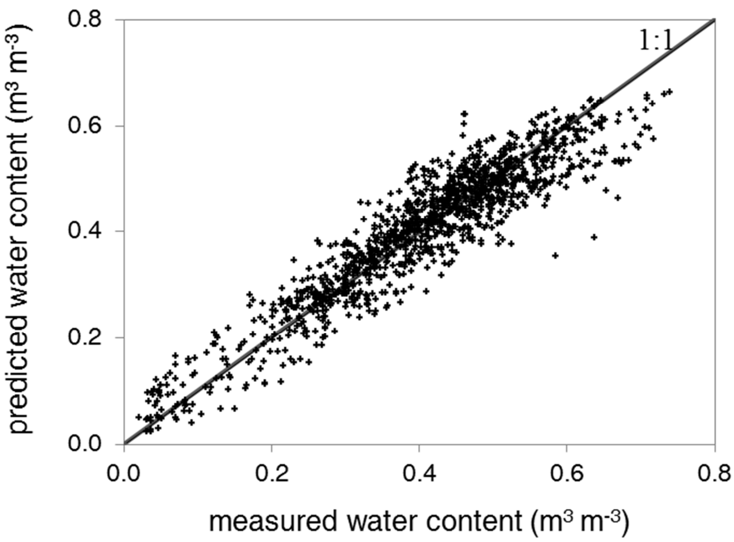

Figure 4-1. Scatter plot of measured water content (m^3 m^{-3}) versus PTFs' predicted water content (m^3 m^{-3}).

4.3.2. Improved effects of grouping by categorical soil structural information on PTFs' accuracy.

As already mentioned, due to specific soil and water management practices for paddy rice cultivation, typical paddy soils tremendously differ from soils of other land use types and management practices in terms of physical and hydrologic characteristics. Soil structure which describes the partial arrangement of structural units and pore system architecture has been proven as an important factor affecting soil water retention characteristics (Pachepsky and Rawls, 2003). The reason for using only the aspect of presence and absence of pedality as grouping criterion was that further details description of soil structural information such as the degree of soil structural development and the shape and size of soil structural units (i.e. soil aggregates) are only considered in the structured soil group, while massive and structureless soil groups do not contain such information. The sample size of structured soils group is rather small (N = 46) and, therefore, taking into consideration these aspects will lead to more detailed subgroups with just a few samples. These minority subgroups are neither suitable nor representative to be used as a training data set for developing soil water retention PTFs of such specific soils.

The influential role of other grouping criteria (e.g., land use types, soil horizons, soil texture) on the PTF's performance and accuracy was also evaluated. The results (not shown) revealed that in most of the cases, grouping criteria could improve the PTF's accuracy to a certain degree at some matric potentials, e.g., RMSE in the range of 0.04 – 0.054 m^3 m^{-3} with mean equals 0.05 m^3 m^{-3} for PTFs developed by the whole data set, and 0.041 - 0.054 m^3 m^{-3} (mean = 0.049 m^3 m^{-3}), 0.041 - 0.053 m^3 m^{-3} (mean = 0.049 m^3 m^{-3}), 0.039 - 0.053 m^3 m^{-3} (mean = 0.049 m^3 m^{-3}) for PTFs developed based on grouping criteria of land use types, soil horizons, and soil texture, respectively. Soil structure information regarding the aspect of presence or absence of pedality performed better in terms of improving PTFs accuracy (RMSE in the range of 0.037-0.054 m^3 m^{-3} with mean equals 0.048 m^3 m^{-3}). This finding is well illustrated by the study of Williams et al. (1983). They emphasized that the presence of pedality is one of the soil properties consistently correlated with differences between the groups of soils with similar water retention characteristics.

The coefficient of determination, R^2, of the point-functions developed from subsets of structured, structureless and massive soils (**Table 4-3**) are smaller than those derived from the whole data set (**Table 4-2**). The reasons leading to the small R^2 value of subgroups were probably the small sample size of the subsets (Pachepsky and Rawls, 2003; Wösten et al., 2001), and less variables being retained to estimate soil water retention (Kutner et al., 2005). For instance, point PTFs derived from structured and structureless soil groups contain only BD, log(OC), Sa and Cl content as single and combined predictors to estimate SWRC (**Table 4-3**).

Table 4-3. Predictors and coefficients of determination (R^2) of point-PTFs developed for different soil groups. Log(OC) is the logarithmic form of organic carbon content (%), BD is soil bulk density (Mg m^{-3}), Sa, Si, Cl are sand, silt, and clay content (%).

Matric potential (kPa)	Massive soil group (N=91)		Structureless soil group (N=23)		Structured soil group (N=46)	
	Predictor	R^2	Predictor	R^2	Predictor	R^2
-1	log(OC), Cl, BD	0.66	BD	0.56	BD	0.34
-3	log(OC), Cl, BD	0.68	BD	0.46	BD, Cl	0.41
-6	log(OC), Cl, BD	0.69	log(OC)	0.51	BD, Cl	0.42
-10	log(OC), Cl, BD	0.73	log(OC)	0.56	BD, Cl	0.41
-20	log(OC), Cl, BD	0.69	Cl, log(OC)	0.61	log(OC), Cl	0.33
-33	BD, Cl , Si	0.68	Cl	0.47	Cl	0.16
-100	Sa, BD	0.42	Cl	0.44	Sa	0.16
-1500	Cl, Si	0.37	BD, Cl, log(OC)	0.19	Cl	0.29

Salchow et al. (1996) reported similar observations. They used the textural class as indicator variable to partition the whole data set into texture-specific soil groups and then conducted regression on data within subsets. They got smaller R^2 values in within-subset regression equations, but the correlation between predicted and observed dependent variables was improved in all cases.

These fairly small R^2 values in the present study are therefore still acceptable because their corresponding functions provide an adequate fit for the data set, and offer the simplest models that can satisfactorily account for the variation in prediction (principle of parsimony, Minasny and Hartemink (2011)).

Notwithstanding the small R^2 values, the point PTFs developed for the three structural soil groups are more accurate in estimating soil water retention as compared with those developed for the whole data set. The average values of RMSE across different pressure heads of SWRC of PTFs derived for structural subgroups are lower than those derived from the whole dataset, especially at high matric potentials (low suctions) (**Figure 4-2**). As it could be expected, the influence of soil structure and its natural pore size distribution on soil water retention is proven to be more robust in the wet moisture range (Or and Wraith, 2002).

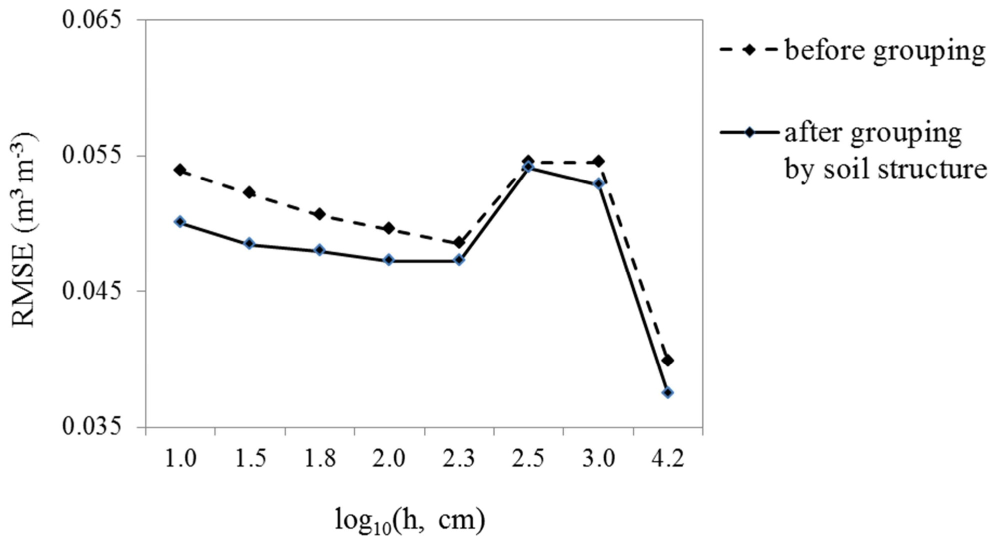

Figure 4-2. Variation of average RMSE values as a function of $\log_{10}h$ (h is the pressure head expressed in cm water) for point PTFs developed before and after grouping by using descriptive soil structural information.

The highest RMSE values were in the range of pF 2.5 – 3.0 (pF is the decimal logarithm of pressure head, h, expressed in cm water; in this range, h varies from -33 kPa, or 336 cm water, to -100 kPa, or 1020 cm water) in both PTFs developed before and after grouping. This finding is supported by Rajkai and Várallyay (1992), who reported lowest PTF accuracy somewhere between -10 to -100 kPa. Moreover Cornelis et al. (2001); Obalum and Obi (2012) also indicated that the prediction error for soil water retention is usually large at high and intermediate matric potentials. In fact, water retention in the wet moisture range is primarily determined by soil structure, which is to a lesser extent related to basic soil properties as it is also greatly affected by several external factors (Kay and Angers, 2002) which might not be accounted for in PTFs as predictors. On the other hand, most accurate estimations are generally obtained in the dry moisture range. The lowest RMSE value at low matric potentials is probably due to the inherent low water content retained in the soil which leads to lesser variation between measurement and prediction compared with the wet and intermediate ranges (Nemes et al., 2006a). The improved accuracy of the PTFs developed based on specific-structure soil groups in our study confirmed the identity of SWRC of samples having similar soil structural morphology (Wösten et al., 2001).

4.3.3. Effects of supplementary predictors in PTF's predictability.

One of the limitations of PTF performance and accuracy which has been reported so far concerns to the lack of sufficient information explaining soil water properties (Nemes et al., 2003). Therefore, beside taking into account the effects of soil structural information in improving PTF's predictability (as presented in the previous section), we also experimented with an extended set of predictors that are widely available or easily measurable, such as pH, EC, soil aggregate stability and plastic limit.

The list of predictors selected from the extended set of predictor variables and the corresponding R^2 values of point PTFs developed before and after grouping are summarized in **Table 4-4**.

Incorporating supplementary predictors slightly increased R^2 values of regression equations (**Table 4-4**) as compared with those obtained using the classical predictor set (**Table 4-3**), especially at high matric potentials (low suctions). For the whole data set, plastic limit together with BD, OC content, and particle size distribution explain more than 72 % of the variability of soil water content at high matric potentials ranging from -1 kPa to -10 kPa. Meanwhile, using soil structural information to partition the data set into three structural subsets, plastic limit solely acts as predictor in characterizing soil water retention at the matric potential range of -1 to -10 kPa for the structured soils group. The ability of the soil to exhibit plastic behavior has been proven to be related to clay and organic matter content (Horn and Baumgartl, 1999; Keller and Dexter, 2012), thus, soil plasticity almost provides the same information in explaining the variability of soil water retention characteristics as soil texture and organic matter do. Therefore, the plastic limit can effectively substitute these two basic properties in estimating soil water retention of soils that exhibit a given degree of soil structure development. The result of the present study is in good accordance with those reported by Khlosi et al. (2013). They found that plastic limit together with other soil properties (e.g., soil texture, soil carbonate content and specific surface area) have a distinct influence on SWRC of Syrian soils.

Table 4-4. Selected predictors and corresponding R^2 values for point-PTFs developed using the extended set of predictors.

Matric potential (cm)	Before grouping (N=160)		After grouping					
			Massive soils (N = 91)		Structureless soils (N = 23)		Structured soils (N =46)	
	Predictor	R^2	Predictor	R^2	Predictor	R^2	Predictor	R^2
-10	PL, BD, Cl	0.72	log(OC), Cl, BD	0.66	BD	0.56	PL	0.55
-30	PL, BD, Cl, log(OC)	0.76	log(OC), Cl, BD	0.67	BD	0.46	PL	0.55
-60	PL, Cl, log(OC)	0.80	log(OC), Cl, BD	0.69	log(OC), SI	0.65	PL	0.52
-100	PL, Cl, log(OC), Si	0.83	log(OC), Cl, BD	0.73	log(OC), SI	0.73	PL	0.46
-200	Sa, log(OC), Cl, BD	0.89	Cl, BD, pH, Si	0.67	log(OC), SI	0.64	Cl, log(OC)	0.33
-340	Sa, BD, Cl	0.83	Sa, BD, Cl, pH	0.68	log(OC), SI	0.53	Cl	0.16
-1.020	Sa, BD	0.80	Sa, BD	0.42	Cl	0.44	Sa	0.16
-15.000	Sa, Cl, BD	0.82	Cl, Si	0.37	pH, EC, SI, Cl, Si, BD, log(OC)	0.38	Cl, EC, PL, log(OC)	0.48

For the structureless soils group, the aggregate stability index (SI), in conjunction with organic carbon content, accounted for 53% to 73% of the variability in soil water retained within the matric potential range of -6 to -33 kPa. However, all predictor variables were needed to significantly explain the variability of soil water content at permanent wilting point (h = -1500 kPa). This is probably due to the wide range of variability of soil water content at this tension in the subset which contained a limited number of observations (N = 43).

In brief, when incorporating supplementary predictors to our derived PTFs, R^2 is slightly increased (**Table 4-4** *vs*. **Table 4-3**). In terms of RMSE, only a very small reduction could be observed (**Figure 4-3**).

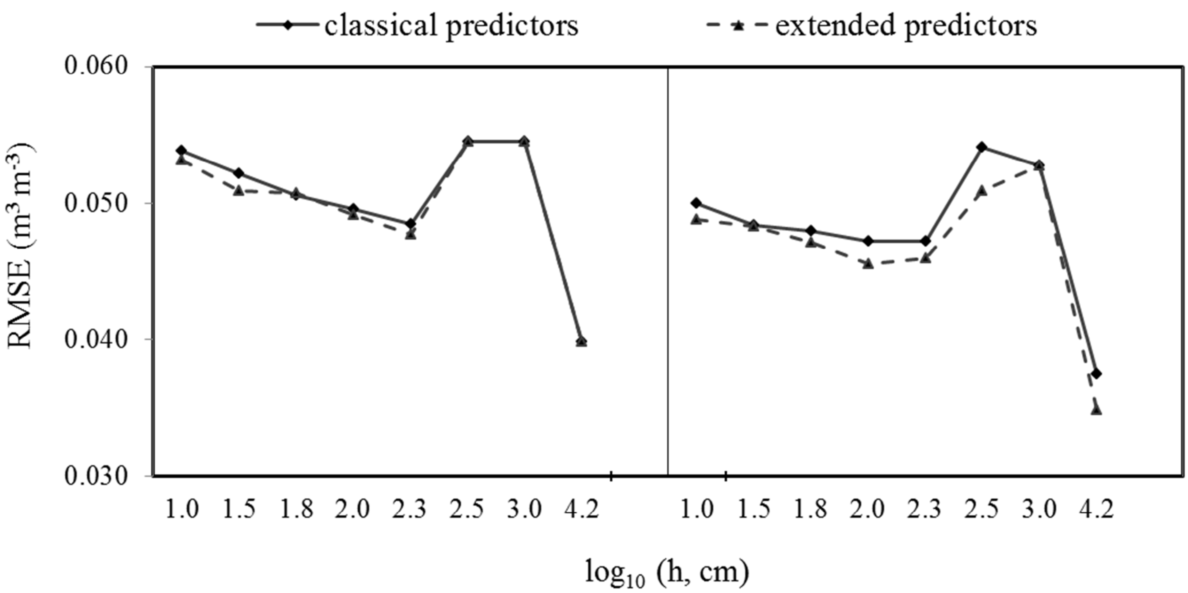

Figure 4-3. Variation of average RMSE values as a function of $\log_{10}h$ (h is the pressure head expressed in cm water) for point-PTFs developed before (A) and after grouping by soil structure (B) using classical predictors set and extended predictors set.

These findings confirm the importance of widely-used water retention PTF predictors such as OC content, soil texture and BD in explaining the variability of SWRC (see e.g., the studies of Gupta and Larson (1979); (Rawls et al., 2003); Shwetha and Varija (2013)), and attaining satisfactorily accurate estimations of soil water retention.

Nonetheless, it is worth mentioning that a positive correlation was found between plastic limit and SWRC for the Mekong Delta soils. This finding is significant because the determination of plastic limit is very simple and does not require any specialized devices. Thus exploiting this property as a potential predictor could be valuable in the development and application of hydraulic pedotransfer functions in tropical developing countries.

4.4. Conclusions

The present study proposes highly applicable point PTFs for estimating soil water retention characteristics of young and fertile tropical delta soils, which are mainly used for paddy rice cultivation. Including categorical soil structural description on top of widely used predictors (e.g. soil texture, bulk density and organic carbon content) as grouping criterion could improve the predictability of SWRC-PTFs. Grouping partitions the soil data set into groups of more uniform soils, thus resulting in more accurate predictive functions. The effect of soil structure on SWRC was most pronounced in the wet moisture range as could be expected from theory. Moreover, as a consequence of the lower deviation within each group, the PTFs developed for each group of soils exploit lesser soil property variables as crucial predictors.

Additionally, the evaluation of supplementary predictors has revealed that plastic limit is a potential predictor for estimating soil water retained at high matric potentials (low suctions) for soils which present a given degree of soil structural development. Determination of plastic limit is easy and less costly and, therefore, plastic limit is a potential predictor which should be taken into consideration when developing hierarchical PTFs based on the availability of soil predictor variables.

One of the drawbacks of our derived point PTFs for specific soil groups is the presence of unexplainable variability of SWRC. The small R^2 value of these PTFs at certain matric potentials for given soil groups might partly result from the small size of the subsets which resulted in high variability within subsets and/or lack of relevant information which impact prediction of soil water retention. When larger data sets will become available in the future, due consideration should be given to the findings from this study to improve the predictability of water retention properties of tropical delta soils.

Chapter 5

COMPARISON OF DIFFERENT REGRESSION TECHNIQUES IN DEVELOPING PTFS

This chapter is rewritten based on:

Phuong Minh Nguyen, Amir Haghverdi, Jan de Pue, Yves-Dady Botula, Khoa Van Le, Willem Waegeman, Wim M. Cornelis (2016). Comparison of statistical regression and data-mining techniques in estimating soil water retention of tropical delta soils. *Biosystem Engineering* (in press).

5.1. Introduction

Although substantial studies have been devoted to develop and evaluate PTFs, several questions still are to be addressed particularly for paddy soils in the tropical delta where the interrelationship between soil and water has not been well established (Pachepsky et al., 2015). One such question relates to the most optimal structural dependency between basic soil properties and soil water retention characteristics (SWRC), which could be formulated by various regression methods. There are two main categories of regression methods which are widely used for PTF development: statistical regression techniques and data mining or pattern-recognition techniques (Pachepsky and Rawls, 2004; Vereecken et al., 2010).

Regarding the state-of-the-art of SWRC-PTFs, most PTFs derived during the past decades are based on statistical regression methods in which the relationship between the basic soil properties and SWRC are quantified by predefined mathematical equations (e.g., the PTFs of (Gupta and Larson, 1979; Hodnett and Tomasella, 2002; Minasny and Hartemink, 2011; Saxton and Rawls, 2006). Recently, alternative data mining techniques such as Artificial Neural Networks (ANN), k-Nearest Neighbors (kNN), and Support Vector Machines for Regression (SVR) have been introduced as promising tools for PTF development (Botula et al., 2014). Firstly, these techniques have been successfully used for both classification and regression problems in other fields of hydrology. For examples, ANN, SVR and kNN techniques were effectively used to forecast rainfall (Hong and Pai, 2007; Hu et al., 2011), water evaporation from soil and free water surfaces (Baydaroğlu and Koçak, 2014), and inflow of water reservoir (Valipour et al., 2012, 2013). Due to their high flexibility and accurate predictive performance, these data mining techniques have gained popularity in unsaturated soil hydrological studies (Botula et al., 2013). These methods have been intensively tested with soils in temperate regions (Lamorski et al., 2008; Nemes et al., 2006a; Pachepsky et al., 1996b; Schaap and Leij, 1998; Twarakavi et al., 2009), and only one kNN study was devoted to highly weathered soils in the humid tropics (Botula et al., 2013). All mentioned authors have confirmed the superiority of the used data mining techniques in modeling the interaction of soil and water as a very complex system compared to traditional linear regression techniques, although several drawbacks have also been noticed in the same time such as susceptibility to over-fitting, highly data-demanding, and expert knowledge requirement.

In the meantime, Pachepsky et al. (2013) have noted that the successfulness of certain regression techniques in terms of providing accurate estimations of SWRC is somewhat controlled by type of PTFs, availability of soil variables used in predictive functions, and size and properties of training databases. Indeed, the data used for calibrating/training the PTFs should account for most of the variation that is likely to be encountered in the area where the data are meant to be used, hence large databases of good quality are generally expected for PTF development (Wösten et al., 2001). This requirement, however, is hard to be fulfilled in many developing countries in the tropic, where just a few extensive soil and water studies have been done so far. Mayr and Jarvis (1999) also reported that using a small set of relevant data, if available, is better than using a large and general data set.

Concerning to the PTF's types that are frequently used to estimate SWRC in the literature, three broad groups have been noticed. They are (1) point-based PTFs that predict the water content at specific chosen matric potentials, (2) parameter-based PTFs that estimate the parameters of analytical expressions of the SWRC, e.g. those of Brooks and Corey (1964); Campbell (1974); van Genuchten (1980), and (3) physical–conceptual PTFs that predict soil hydraulic properties based on a soil structural model (Cornelis et al., 2001; Wösten et al., 2001). The latter has not been widely used in practice due to some limitations mentioned in Cornelis et al. (2001).

For modeling purposes, the parameter-based PTFs which offer the prediction of a whole and continuous SWRC are often preferred, since many flux transport models require the complete SWRC as input parameters (Cornelis et al., 2001). However, using statistical validation analysis, many researchers (Merdun et al., 2006; Pachepsky et al., 1996a; Tomasella et al., 2003; Vereecken et al., 2010) have noted that point-based PTFs outperformed the parameter-based PTFs in predicting soil water content. The supremacy of point-PTFs could be attributed to the fact that soil water retention at specific matric potentials is controlled by different basic soil properties (Tomasella et al., 2003; Vereecken et al., 2010), and therefore, point PTFs should provide a better combination of these properties and lead to more accurate functions for SWRC estimation. Recently, Haghverdi et al. (2012)

introduced a new PTF approach, named “pseudo continuous” PTFs (PC-PTFs), which are capable to determine almost continuous SWRCs without using any analytical soil hydraulic expressions. In their PTF, using matric potentials as predictor variable enables to predict the corresponding water content at any desired matric potential. They proved that PC-PTFs derived by the ANN technique were more accurate and reliable than parameter-based PTFs, and slightly better than point-PTFs when a limited data set was available for PTFs’ development.

Moreover, due to the very specific nature in terms of physical and hydraulic soil characteristics of tropical delta soils where the main agricultural practice is paddy rice cultivation (**Chapter 4**), it is not advisable to utilize PTFs reported so far in the literature to estimate the soil water characteristics in the tropical delta region (**Chapter 3**). Therefore, the objectives of this study were (1) to develop and validate point PTFs and PC-PTFs using available limited data sets from the tropical Vietnamese Mekong delta, and (2) to investigate the predictive capability of various regression techniques (i.e., MLR, ANN, SVR and kNN) in estimating SWRC in both point and pseudo-continuous manners. Since the concept of PC-PTFs has only been tested and compared against ANN and SVM techniques for soils in dry and temperate regions (Haghverdi et al., 2012; Haghverdi et al., 2014), we believe that testing the predictive power of these approaches together with others as MLR and kNN would be very useful to those that are in need of SWRC data and/or attempt to develop new PTFs for soils in tropical humid region. To our knowledge, this is the first study considering data mining techniques for developing point and PC-PTFs of a variety of soils in a tropical delta dominated by rice paddy cultivation.

5.2. Materials and Methods

5.2.1. Soil data sets

Two data sets collected from VMD were employed to calibrate and validate the point and PC-PTFs in this chapter.

The first data set of 160 samples was used to develop the PTFs (so-called training data set). After the training phase, the performance of derived PTFs was validated by an independent data set of 29 samples taken from 10 soil profiles within the same study region. Detail information about the distribution of soil physical, chemical, and hydraulic properties of the two data sets were summarized in **Chapter 2**.

5.2.2. Types of soil water retention PTFs

Two types of PTFs were derived to estimate SWRC using various regression techniques (which will be described hereinafter). The first one pertains to point PTFs that estimate soil water content at soil matric potentials of –1, –3, –6, –10, –20, –33, –100, –1500 kPa. The inputs used are basic soil properties which have been widely used for SWRC estimations (i.e., sand, silt, clay content, bulk density, and organic carbon content). The second type refers to PC-PTF which uses the logarithm of matric heads as extra input variable, hence supposedly allowing the prediction of soil water content at any desired matric potentials. The structural topologies of the point-based, and pseudo-continuous PTFs are manifested in **Figure 5-1**.

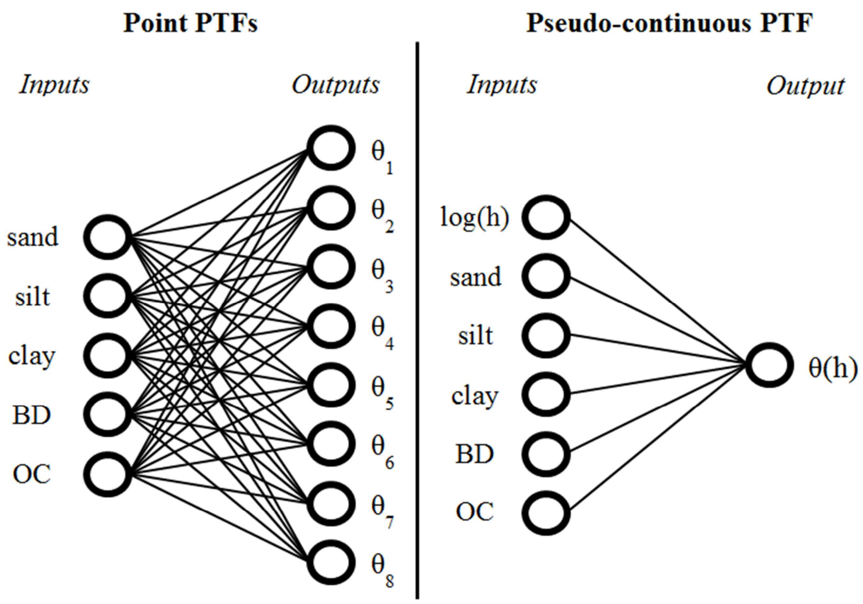

Figure 5-1. Structural topologies of point PTFs, and pseudo-continuous PTFs (modified from Haghverdi et al., 2012)

As can be observed in **Figure 5-1**, sand, silt and clay, BD and OC content are the common input predictors of SWRC-PTFs. θ_1 to θ_8 are volumetric water contents which in turn are the outputs of the point PTFs. In case of PC-PTFs, h (in the unit of cm water) is matric head and is the extra input variable. $\theta_{(h)}$, the output of the PC-PTFs, is volumetric water content at h matric head. Different h values yield different water contents.

5.2.3. Methods to derive PTFs

5.2.3.1. Multiple Linear Regression (MLR)

The general form of MLR-based PTF is as follows:

$$Y_i = aX_1 + bX_2 + cX_3 + \ldots + zX_n \tag{5-1}$$

where Y_i denotes the response variable (for point-PTFs: Y_i with i = 1, 2,…, 8 is the water content at 8 matric potentials, while in case of PC-PTF, Y is the water content at any corresponding matric head input), $X_1, X_2, X_3, \ldots X_n$ are predictor variables (i.e., sand, silt, clay content, bulk density and organic matter content in case of point PTFs, and all aforementioned variables plus logarithm of matric head for the PC-PTF), and *a, b, c, … z* are the values of the regression coefficients obtained from fitting the equation to the training data set. Further detailed explanation about the theory of the MLR method can be found in Kutner et al. (2005).

The R statistical language (R Core Team, 2014) was used to develop the MLR-based PTFs, in which the potential and significant predictors for SWRC estimation were selected by stepwise regression using Akaike Information Criteria (AIC) as selection criterion. Akaike Information Criterion is a measure of the relative quality of a statistical model for a given set of data. AIC deals with trade-off between the goodness of fit of the model and the complexity of the model.

5.2.3.2. Artificial Neural Networks (ANN)

The point and pseudo-continuous ANN-PTFs were established in Matlab R2014a environment (the MathWorks). A three-layer feed forward back propagation ANN model was selected. The activation functions were sigmoid tangent hyperbolic and linear in hidden and output layers, respectively. The Levenberg-Marquardt algorithm (Demuth and Beale, 2000) was implemented for training. The number of neurons in the hidden layer was changed from 1 to 20. The input and outputs of the ANN-PTFs were identical to those of MLR-PTFs. The bootstrap method (Efron and Tibshirani, 1993) was applied on training data to create 50 statistically similar subsets of the same size through a sampling with replacement technique. The subsets, comprising about 63% of the parent data (Schaap et al., 2001), were used to train the PTFs. The idle samples (i.e. 37%) formed the cross-validation set. The training was stopped whenever the error increased on the cross-validation set. The outputs of the 50 bootstraps were averaged and reported as the predictions from the PTFs.

5.2.3.3. Support Vector Machines for Regression (SVR)

The SVM algorithm was implemented in the R statistical language (R Core Team, 2014) to derive point and PC-PTFs. The most commonly used kernel, i.e. radial basis function kernel, which has been applied in the works of Haghverdi et al. (2014); Lamorski et al. (2008); Twarakavi et al. (2009) was selected to build our SVR-models. The optimal hyper-parameters of the SVR models were estimated using a thorough grid-based search approach (Hastie et al., 2009) in which the parameter C was changed from 0.001 to 1 in increment of 0.1, ε was varied from 0 to 1 in increment of 0.05, while γ was adjusted from 0.01 to 1 with a mesh increment of 0.1. The range of the meta-parameters was obtained from the preliminary grid searches with large mesh of increment. A two-round grid search was implemented due to the high computational cost of the SVR optimization process for three parameters simultaneously. Ten-fold cross validation technique was used in the grid-based search; and the set of C, γ, ε corresponding to the best cross-validation accuracy was picked and used to calibrate SVR models.

Detailed information about the methodology of this technique can be found in many previous works, e.g., Lamorski et al. (2008); Smola and Schölkopf (2004); Twarakavi et al. (2009).

5.2.3.4. k-Nearest Neighbors (kNN)

The basic idea of the kNN technique, named similarity-based technique by Nemes et al. (2006a), is finding the k nearest neighbors from a reference dataset for each soil in the test dataset in terms of selected input attributes. For kNN estimation, two design parameters were defined and used for the estimation procedure, namely the k and p terms. The k term refers to the number of similar soils to be selected from the reference data set 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. Working with extensive soil databases in temperate regions, Nemes et al. (2006a) derived regression equations relating the *k* and *p* terms with training data set size. Botula et al. (2013) tested this relation for soils in the humid tropics and obtained similar results. Therefore, in this study, we use the proposed formula of Nemes et al. (2006a) for determining the designed parameters *k* and *p*.

More methodological and calculation details on the whole procedure can be found in the works of Botula et al. (2013); Nemes et al. (2006a); Nemes et al. (2006b). The kNN algorithm used in this study was adapted from the variants developed by Nemes et al. (2006a) and Botula et al. (2013). The implementation of the kNN algorithm was done in the Matlab R2014a environment (the MathWorks).

5.2.4. Evaluation criteria

The evaluation of model performance (i.e., accuracy and reliability) is commonly made and reported through the comparison of PTF estimated and observed values. As defined by Wösten et al. (2001), accuracy of a PTF is the correspondence between measured and predicted data for the data set from which a PTF has been developed; whereas reliability of PTF is the correspondence between measured and predicted data for the data set other than the one used to develop a PTF. In this chapter, three statistical indices, i.e. (1) mean of prediction error (ME), a measure of the prediction bias which indicates the over- or under-estimations of a specific model, (2) the root mean square of the prediction error (RMSE), a measure of the overall prediction error, and (3) the coefficient of determination (R^2) which indicates the amount of variation in the data explained by the regression model, were selected to assess the predictive ability of the derived PTFs in both calibration and validation phases. These statistical indices were calculated using the equations in **Chapter 3.**

For the MLR and SVR approaches, all soils in the training data set were used to evaluate the accuracy of the derived PTFs (N is equal to 160 and 1280 observations for point and PC-PTFs, respectively). For the kNN algorithm, the accuracy of the models was evaluated based on the leave-one-out (LOO) procedure as it is an instance-based regression method which needs separate test data for evaluation. As the name suggests, during the evaluation procedure, one sample was left out and the remaining samples were used as the training data to derive the estimation for the leave-out sample (Mucherino et al., 2009). Therefore, N=160-

1=159 observations for point estimation. In case of PC_{kNN}-PTF, eight corresponding water retention points of the leave-out soil were removed, hence N=1280-8=1272 observations for pseudo-continuous estimation. The reliability of point PTFs derived by these regression methods was assessed by using the test data set of N = 29 samples, corresponding to N = 29 x 8 = 232 samples for PC-PTFs. More details about “accuracy” and “reliability” terminologies can be found in Wösten et al. (2001).

5.3. Results

5.3.1. Exploratory data analysis

In order to find interrelations of soil properties in the training data set, an exploratory data analysis was conducted. The Pearson correlation matrix (**Table 5-1**) displayed significant correlation between soil moisture retained at different matric potentials and particle size distribution (i.e., sand, silt, clay content), BD, and OC content. Clay, silt and OC content were positively correlated with soil water content, whereas BD and sand content were negatively correlated. The correlation strength between OC or BD, and soil moisture increased with increasing soil matric potentials (i.e., less negative), whereas those of clay or sand content rose with decreasing matric potentials. These observations were expected because soil structure which determines water content stored in soils at high potentials is more related to organic carbon content and BD (Botula et al., 2013; Pachepsky et al., 2006), whereas soil water content at the dry end of the SWRC is primarily determined by adsorption forces in the soil matrix (mainly determined by soil texture, particularly clay content) (Manrique et al., 1991). Additionally, there is strong correlation between OC content and soil bulk density (r = -0.75), confirming that soils with higher organic matter will concomitantly have lower bulk density. On the other hand, significant but weak correlations were observed between clay, sand contents and OC. This is most probably because the training data were collected in the region where paddy rice is the main agricultural practice. Rice soils, coincident with fine-textured soils, often have high OC accumulation in the surface due to long-lasting submerged condition of paddy-rice cultivation (Linh et al., 2015a).

The exploratory data analysis also exposed an exponential relationship between total OC content and soil moisture content retained at different pressure heads in both data sets. In order to properly deriving optimal PTFs based on the linear regression technique, a log-transformation was applied to resolve the non-linearity problem of total OC content before conducting MLR analysis. Data mining techniques, on the other hand, can handle highly non-linear data, and therefore original soil variables in the training data set were used to train ANN, SVR and kNN models.

Table 5-1. Pearson's correlation coefficients between soil properties in the training data set (N=160).

Soil properties	clay	silt	sand	BD	OC	logOC	θ_{-1} kPa	θ_{-3} kPa	θ_{-6} kPa	θ_{-10} kPa	θ_{-20} kPa	θ_{-33} kPa	θ_{-100} kPa	θ_{-1500} kPa
clay	1													
silt	0.33**	1												
sand	-0.87**	-0.74**	1											
BD	-0.47**	-0.23**	0.45**	1										
OC	0.35**	0.07	-0.28**	-0.75**	1									
logOC	0.61**	0.33**	-0.6**	-0.80**	1.00**	1								
$\theta_{-1\ kPa}$	0.71**	0.27**	-0.64**	-0.72**	0.65**	0.75**	1							
$\theta_{-3\ kPa}$	0.74**	0.31**	-0.68**	-0.72**	0.64**	0.77**	0.99**	1						
$\theta_{-6\ kPa}$	0.80**	0.41**	-0.77**	-0.68**	0.62**	0.80**	0.94**	0.97**	1					
$\theta_{-10\ kPa}$	0.82**	0.46**	-0.82**	-0.65**	0.59**	0.80**	0.90**	0.93**	0.99**	1				
$\theta_{-20\ kPa}$	0.84**	0.51**	-0.85**	-0.63**	0.54**	0.76**	0.84**	0.88**	0.95**	0.98**	1			
$\theta_{-33\ kPa}$	0.82**	0.53**	-0.86**	-0.61**	0.49**	0.71**	0.79**	0.82**	0.90**	0.94**	0.99**	1		
$\theta_{-100\ kPa}$	0.80**	0.56**	-0.85**	-0.56**	0.40**	0.65**	0.71**	0.75**	0.82**	0.87**	0.94**	0.98**	1	
$\theta_{-1500\ kPa}$	0.84**	0.54**	-0.87**	-0.49**	0.33**	0.60**	0.7**	0.74**	0.81**	0.85**	0.90**	0.93**	0.95**	1

** shows significant correlation at 0.01 significance level.

5.3.2. Point PTFs performance

The R^2, ME, and RMSE of point PTFs derived by different methods are summarized in Table 3 and Fig. 3. All of the four prediction methods displayed a comparable performance in the training phase, in which the ANN-PTFs performed best. The relative high value of coefficient of determination (average R^2 equals 0.79, 0.84, 0.81, and 0.75 for MLR-, ANN-, SVR- and kNN-PTFs, respectively) indicates that a large proportion of the SWRC variability of training samples can be explained by these empirical models using basic soil data (i.e., soil texture, BD and OC content). Similarly, the log(h) dependent analysis of RMSE (left panel of **Figure 5-2**) displays satisfactory accurate estimations of soil water content at different pressure heads with RMSE ranges of 0.040-0.054 $m^3 m^{-3}$ for MLR-PTFs, 0.036-0.045 $m^3 m^{-3}$ for ANN-PTFs, 0.039-0.053 $m^3 m^{-3}$ for SVR-PTFs, and 0.044-0.059 $m^3 m^{-3}$ for kNN-PTFs. The scatter plots of observed *vs.* PTF predicted soil water content of different models (left panel of **Figure 5-3**), together with approximately-closed zero values of ME in the training phase (**Table 5-2**), once again strengthen the appropriateness of these methods in describing SWRC of tropical lowland delta soils. Most of the points are closely scattered around the reference 1:1 lines and do not exhibit much bias.

Further detailed assessment of all error measures (i.e., ME, RMSE and R^2) of these methods (**Table 5-2** and **Figure 5-2**), however, expose a marginally worse performance of kNN-PTFs compared to ANN-, SVR- and MLR-PTFs. Notwithstanding this, kNN as well as other pattern recognition techniques like ANN and SVR were more accurate than MLR models in predicting soil water retention of independent test samples (**Table 5-2** and right panel of **Figure 5-2** and **Figure 5-3**). The average values of RMSE, ME and R^2 for the validation data set equal 0.049 $m^3 m^{-3}$, -0.028 $m^3 m^{-3}$, 0.89 for kNN-PTFs; 0.052 $m^3 m^{-3}$, -0.032 $m^3 m^{-3}$, 0.88 for SVR-PTFs; and 0.053 $m^3 m^{-3}$, -0.034 $m^3 m^{-3}$, 0.82 for ANN-PTFs, respectively. These results asserted that the reliability of point PTFs derived by the data mining or pattern-recognition approaches are much better than that of MLR-PTFs (RMSE = 0.068 $m^3 m^{-3}$, ME = -0.043 $m^3 m^{-3}$, R^2 = 0.84).

Concerning to the variation of RMSE in dependence on the soil matric potentials (right panel of **Figure 5-2**), all four methods exposed a similar trend (i.e, RMSE was relatively low at the wet part of the curve, increased toward the intermediate region, but decreased again at the dry part) with the extreme climaxes observed with MLR-PTFs. Comparable patterns of error variation were also reported by many researchers (Botula et al., 2013; Haghverdi et al., 2015; Vereecken et al., 2010). These authors asserted that the accuracy of PTFs in dependence on the matric potentials are affected by the PTF type, the data characteristic, and the input attributes combination.

Table 5-2. Mean prediction error (ME) and coefficient of determination (R^2) of point PTFs derived based on various regression techniques, i.e., Multiple Linear Regression (MLR), Artificial Neural Networks (ANN), Support Vector Machines for Regression (SVR), and k-Nearest Neighbors (kNN).

Matric potentials	Training phase				*Testing phase*			
	MLR	ANN	SVR	kNN	*MLR*	*ANN*	*SVR*	*kNN*
				ME ($m^3 m^{-3}$)				
-1 kPa	$-8.7*10^{-17}$	0.0013	-0.0004	-0.0002	*-0.040*	*-0.032*	*-0.034*	*-0.028*
-3 kPa	$-1.8*10-^{16}$	0.0014	0.0006	-0.0007	*-0.046*	*-0.036*	*-0.029*	*-0.031*
-6 kPa	$-1.1*10^{-16}$	0.0011	-0.0013	-0.0016	*-0.066*	*-0.043*	*-0.035*	*-0.037*
-10 kPa	$-6.3*10^{-17}$	0.00002	0.0029	-0.0018	*-0.075*	*-0.050*	*-0.043*	*-0.042*
-20 kPa	$-7.6*10^{-17}$	0.0002	-0.0009	-0.0024	*-0.071*	*-0.056*	*-0.054*	*-0.049*
-34 kPa	$-3*10^{-17}$	0.0014	-0.003	-0.0026	*-0.061*	*-0.054*	*-0.056*	*-0.049*
-100 kPa	$-8*10^{-17}$	0.0006	-0.006	-0.0021	*-0.024*	*-0.034*	*-0.040*	*-0.028*
-1500 kPa	$7.4*10^{-16}$	0.0001	-0.0004	-0.0005	*0.043*	*0.033*	*0.035*	*0.041*
Average	$\mathbf{1.4*10^{-17}}$	**0.0008**	**-0.001**	**-0.001**	***-0.043***	***-0.034***	***-0.032***	***-0.028***
				R^2				
-1 kPa	0.71	0.78	0.79	0.68	*0.84*	*0.83*	*0.88*	*0.89*
-3 kPa	0.75	0.81	0.78	0.71	*0.84*	*0.83*	*0.88*	*0.90*
-6 kPa	0.80	0.84	0.79	0.76	*0.79*	*0.83*	*0.92*	*0.90*
-10 kPa	0.83	0.87	0.82	0.79	*0.78*	*0.83*	*0.87*	*0.90*
-20 kPa	0.84	0.88	0.83	0.79	*0.84*	*0.81*	*0.88*	*0.89*
-34 kPa	0.81	0.86	0.81	0.76	*0.87*	*0.82*	*0.88*	*0.89*
-100 kPa	0.77	0.83	0.82	0.73	*0.91*	*0.83*	*0.91*	*0.88*
-1500 kPa	0.79	0.83	0.81	0.76	*0.85*	*0.79*	*0.82*	*0.88*
Average	**0.79**	**0.84**	**0.81**	**0.75**	***0.84***	***0.82***	***0.88***	***0.89***

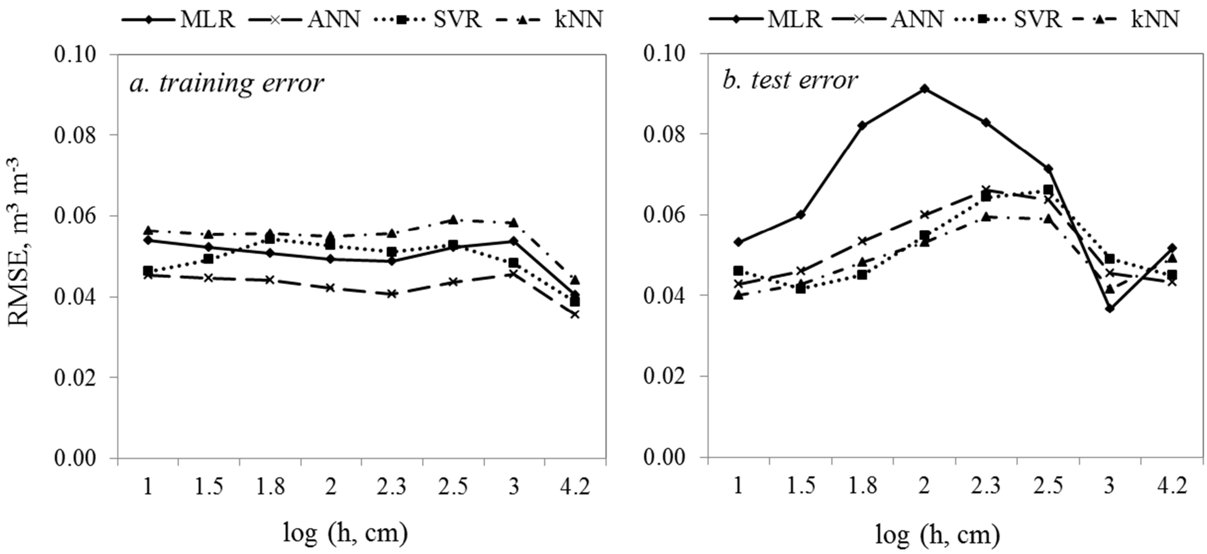

Figure 5-2. Variation of root mean square error (RMSE) as a function of log(h) (h is matric head, expressed in cm water) in training phase and testing phase of point PTFs derived by Multiple Linear Regression (MLR), Artificial Neural Networks (ANN), Support Vector Machines for Regression (SVR), and k-Nearest Neighbors (kNN) methods.

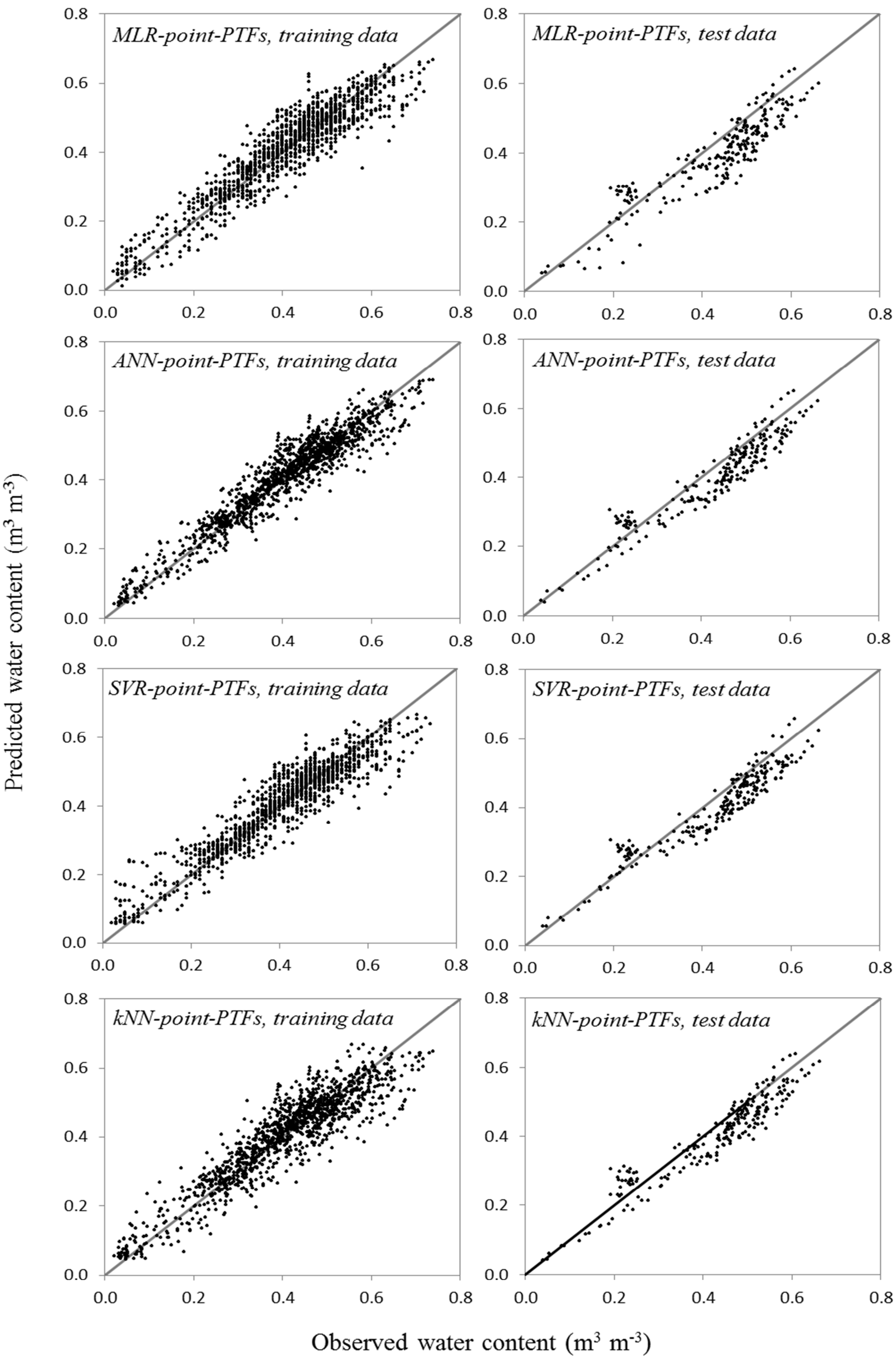

Figure 5-3. Scatter plots of observed vs. predicted water content (m^3 m^{-3}) of point PTFs in calibration and validation phases. MLR-point-PTF, ANN-point-PTF, SVR-point-PTF, kNN-point-PTFs are the point PTFs derived by Multiple Linear Regression, Artificial Neural Networks, Support Vector Regression, and k-Nearest Neighbors techniques.

5.3.3. Pseudo-continuous PTFs (PC-PTFs)

The predictive performance in terms of ME, RMSE and R^2 of PC-PTFs derived by different regression methods is summarized in **Table 5-3**. The evaluation of the accuracy and reliability show that the SVR and MLR techniques are probably not a proper choice of tools for the development of PC-PTFs, at least within the context of this study.

Table 5-3. Performance evaluation using different statistical indices as RMSE, ME and R^2 of point PTFs and PC-PTFs developed by multiple linear regression (MLR), artificial neural networks (ANN), support vector machines for regression (SVR) and k-nearest neighbors (kNN) methods.

Regression methods		RMSE		ME		R^2	
		Point PTFs	**PC-PTFs**	Point PTFs	**PC-PTFs**	Point PTFs	**PC-PTFs**
Accuracy	MLR	0.050	**0.056**	$1.4*10^{-17}$	$\mathbf{-3*10^{-18}}$	0.79	**0.84**
	ANN	0.043	**0.044**	0.0008	**-0.0007**	0.84	**0.90**
	SVR	0.049	**0.036**	-0.001	**-0.0009**	0.81	**0.93**
	kNN	0.055	**0.056**	-0.001	**0.0003**	0.75	**0.84**
Reliability	MLR	0.068	**0.066**	-0.043	**-0.043**	0.84	**0.85**
	ANN	0.053	**0.052**	-0.034	**-0.035**	0.82	**0.90**
	SVR	0.052	**0.068**	-0.032	**-0.044**	0.88	**0.84**
	kNN	0.049	**0.050**	-0.028	**-0.027**	0.89	**0.90**

Using the MLR technique, the accuracy of derived PC-PTF (i.e. in calibration phase) in terms of overall prediction error is worse (RMSE = 0.056 m^3 m^{-3}) than that of point PTFs (RMSE = 0.05 m^3 m^{-3}). The points of the scatter plot of PC_{MLR}-PTFs (**Figure 5-4**) are more dispersed with a sign of underestimation in the wet moisture range and overestimation in the dry moisture range compared to those of point PTFs derived by the same method (**Figure 5-3**). The reliability of PC_{MLR}-PTF in terms of validation error (RMSE = 0.066 m^3 m^{-3}) is comparable with the average value of $point_{MLR}$ PTFs (RMSE = 0.068 m^3 m^{-3})

Regarding SVR, the results show that the PC_{SVR}-PTF derived based on the optimal SVR meta-parameters obtained from the 10-fold cross validation process (i.e., C=1, γ=0.6, ε=0.05) offers a very best fit to the training data set with R^2 = 0.93, ME = -0.0009 m^3 m^{-3}, and RMSE = 0.036 m^3 m^{-3}. The scatter plot of PC_{SVR}-PTF displays a substantial agreement between observed and PC-PTF predicted water content (**Figure 5-4**). However, the validation result of the derived PC_{SVR}-PTF exposes a poor generalization performance to the test samples. The validation error is rather high (RMSE = 0.068 m^3 m^{-3}) and in the same order of magnitude with the one yielded by PC_{MLR}-PTF.

The PC-PTF approach was recommended for limited available training data when developed by highly data-demanding techniques (Haghverdi et al., 2012). The suitability of the ANN method in developing PC-PTFs, which was reported in the studies of Haghverdi et al. (2012); Haghverdi et al. (2014) for soils in dry regions, was confirmed in this study for soils in the tropical humid delta in both training (R^2 = 0.90, ME = -0.0007 m^3 m^{-3}, and RMSE = 0.044 m^3 m^{-3}) and testing phase (R^2 = 0.90, ME = -0.035 m^3 m^{-3}, and RMSE = 0.052 m^3 m^{-3}).

Regarding the kNN method, which was actually applied in this study for the first time to develop PC-PTF, the accuracy of PC_{kNN}-PTF (RMSE = 0.056 m^3 m^{-3}) was comparable to that of $point_{kNN}$ PTFs (RMSE = 0.055 m^3 m^{-3}). A similar agreement in terms of R^2 and graphical correspondence between measured and predicted water content was also observed (**Table 5-3** and **Figure 5-4**). The reliability of such PC_{kNN}-PTF on the test samples was as good as the $point_{kNN}$ PTFs (RMSE = 0.05 m^3 m^{-3}, and R^2 = 0.9).

Regardless of PTF types and regression methods, all derived PTFs underestimate the soil water content of the test samples (ME<0; right panel in **Figure 5-3** and **Figure 5-4**). Although soils in the test data set came from the same population as the training data, the bias of the estimation might probably be due to either spatial variability (i.e., sampling sites of training and testing data sets), or temporal variation of soil hydraulic characteristics as a result of changes in land use types and soil management (e.g., manifested by the difference in the range of OC content between two data sets) (Or and Wraith, 2002). Nonetheless, the predicted SWRC from outperforming PTFs (i.e., PTFs derived by SVR for point estimation, ANN and kNN for both point and pseudo-continuous estimation) are of acceptable accuracy for indirect estimation approaches, as these RMSE values (Table 4) are in the typical RMSE ranges reported by Vereecken et al. (2010).

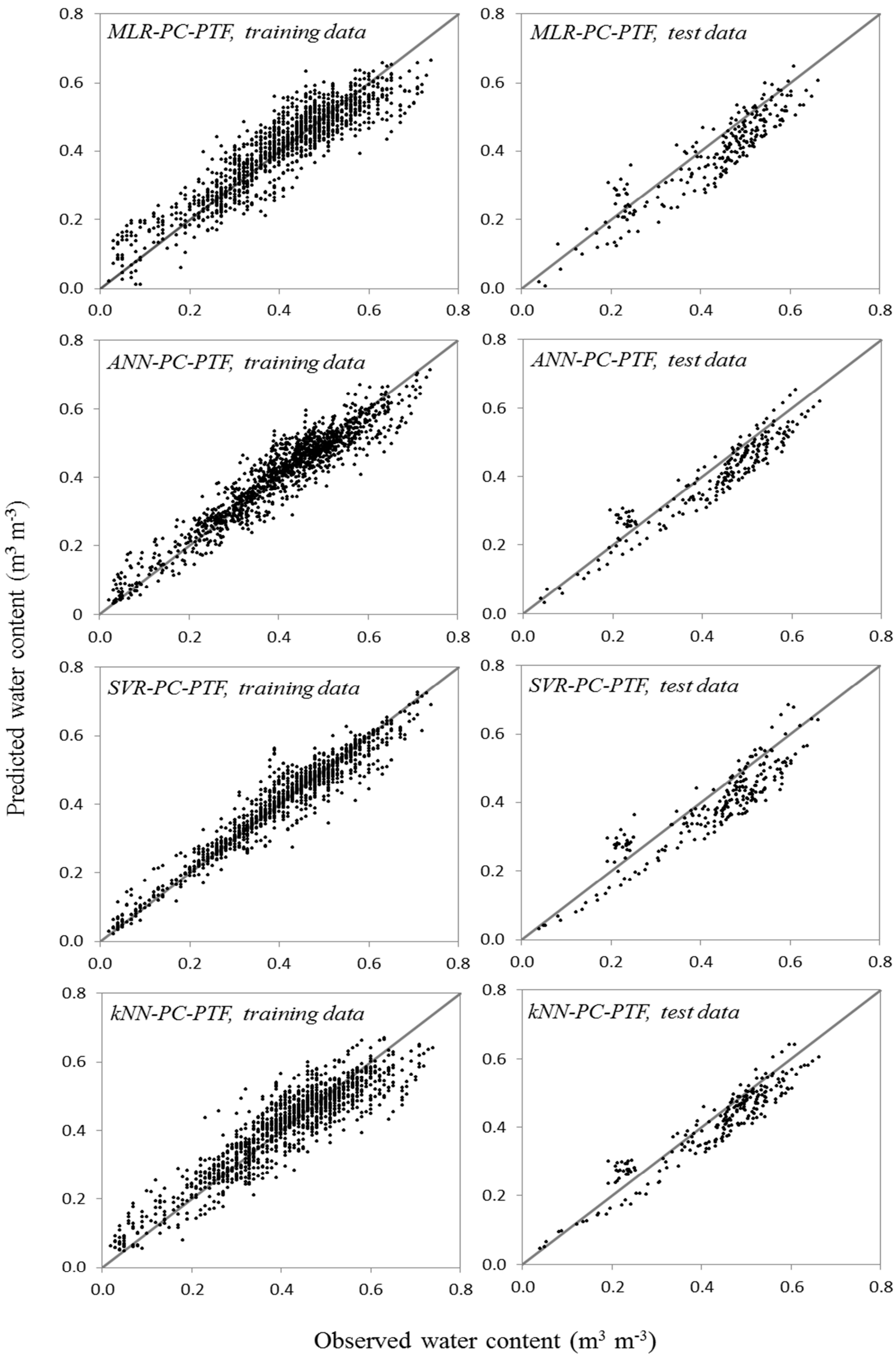

Figure 5-3. Scatter plots of observed vs. predicted water content (m^3 m^{-3}) of pseudo-continuous PTFs in calibration and validation phases. MLR-PC-PTF, ANN-PC-PTF, SVR-PC-PTF, kNN-PC-PTFs are the pseudo-continuous PTFs derived by Multiple Linear Regression, Artificial Neural Networks, Support Vector Regression, and k-Nearest Neighbors techniques.

5.4. Discussions

The generalization strength of data mining techniques (i.e., ANN, kNN and SVM) reported in previous studies, e.g., Botula et al. (2013); Haghverdi et al. (2012); Patil et al. (2013); Twarakavi et al. (2009) was confirmed in this study for point estimation of SWRC of tropical delta soils. Indeed, the pattern recognition techniques of SVR, kNN and ANN do not appear to rely on any stringent assumption about the underlying data and can adapt to any situation, hence providing flexible and reliable estimation (Hastie et al., 2009).

Good generalization performance of the point PTFs derived by the SVR technique might probably result from the implementation of a structural risk minimization in the optimization algorithm. This aspect leads to a better generalization capacity of SVR models for new samples as compared to the statistical linear regression models which employed only empirical risk minimization. Also, the satisfactory generalization performance of kNN-PTFs in point estimation of SWRC would possibly be explained by the similarity-based nature of kNN model together with the synchronous SWRC of soils in both training and testing data sets. Indeed, Perkins and Nimmo (2009) and Botula et al. (2013) have stressed that the predictive capability of PTFs derived by pattern recognition techniques depend on the quality and the level of representability of the training data set to soils for which one needs to predict SWRC. The excellent ability of the ANN technique to mimic the inputs-outputs relationship of complex soil water system (Pachepsky and Schaap, 2004) might probably clarify the adequate performance of ANN-PTFs in both training and testing phases of point and pseudo-continuous estimation. Inversely, the MLR models were constructed based on rigorous structural assumptions of the relationship between SWRC and other soil variables. Hence, the regression equations yield stable but possibly inaccurate estimation (Hastie et al., 2009). It is manifested by poor results of point$_{MLR}$ PTFs with test data in this study.

It is important, however, not to overemphasize the generalization performance of data mining techniques. The prediction capacity of PC-PTFs derived by SVR method for new test samples is comparable with that of statistical linear regression models. The poor generalization performance of PC_{SVR}-PTF in this study opposes to the strengths of SVM algorithms (i.e., promising generalization performance and capacity to handle with non-linear data), which have been reclaimed by other researchers (Lamorski et al., 2008; Twarakavi et al., 2009) for other PTF types (i.e., point-PTFs and parameter-based PTFs). The counter result in present study is supported by the study of Haghverdi et al. (2014). They noted that for the application of the SVM technique, using only the statistical mean square error (MSE) index to select the optimal model is insufficient for PC-PTF type, because the models that show satisfactory values of mean square error in the training phase are the ones displaying a linear relationship between soil-water content and the logarithm of soil matric head. The reliability of PC_{SVR}-PTF is therefore similar to that of PC_{MLR}-PTF.

Although the kNN models provide better estimation of both single and multiple points of SWRC than MLR and SVR models, it is important to note the limitation of the kNN method in developing PC-PTF. Unlike other parametric regression techniques (i.e., MLR, ANN, SVM) which define relationships of soil properties under mathematical functions, the kNN is a non-parametric regression technique which has limitation to provide a continuous prediction of SWRC. Moreover, as the concept of PC-PTFs in combination with the kNN technique was applied for the first time in this study, we would like to clarify the difference of this PC_{kNN}-PTF from that of Nemes et al. (2006a). Since the pseudo-continuous topology considers matric potential as extra input variable, the PC-PTF could consider multiple water retention points of a particular soil in the reference/training data which has basic soil properties very similar to the target soils as nearest neighbors. Such selection opposes to the classical kNN-PTF for point estimation (i.e., selecting multiple soils). Therefore, it should actually be considered as point PTF with matric head as additional input variable for the estimation of several points of SWRC. A continuous SWRC can then be obtained by fitting analytical equations to multiple predicted water retention points. Recently, Haghverdi et al. (2015) have introduced kNN-VG-PTFs in which a non-parametric kNN technique was applied in combination with the van Genuchten model. By this way, any points of SWRC of the target sample could be estimated based on the VG parameters of nearest samples withdrawn from the reference database. Such PTFs showed reasonable accuracy and reliability in comparison with other well-known parameter based PTFs. However, as we already presented in the introduction section, this type of PTFs is beyond the scope and was not tested in this study.

It is worthy to notice the important sign provided by the scatter plots of both PTFs types (point *vs.* PC) derived by various regression methods in the training phase. The fact that the tip of the data plumes (at high matric potentials) is mostly below the 1:1 reference line (left panel of **Figure 5-3** and **Figure 5-4**), might indicate that in this study significant predictors of SWRC might not be optimally identified at high matric potentials. It has been widely shown that soil water retention at high matric potential is primarily determined by the soil-pore system, or in other terms ‘soil structure’. Moreover, in the study of Pulido Moncada (2014), soil structure has been reported as the temporal indicator of the change of soil quality. As it is manifested with the mean test error in the present study, soil temporal variability is probably one of the hindrances of PTF’s transferability. Utilizing soil structure information as one of PTFs predictors is expectedly to improve PTFs’ performance (Vereecken et al., 2010). As it has been noticed in **Chapter 4**, incorporating categorical soil structure information in point PTFs developed by the MLR technique improved the accuracy of the SWRC estimation of tropical paddy soils. It is interesting to further investigate whether such improved effect will still be captured by different data mining techniques and for other PTF types.

5.5. Conclusions

This study presented the development and validation of point and PC-PTFs to estimate SWRC of tropical delta soils from basic soil properties using various regression techniques like MLR, ANN, SVR and kNN. Evaluating the accuracy and reliability of derived PTFs asserted that all four regression techniques provide comparable accuracy in estimation of soil-water content at specific matric potentials, but the reliability of point PTFs derived by pattern recognition techniques of ANN, SVR and kNN was better than that of MLR. In case of PC-PTF, ANN generates more accurate and reliable PTFs than SVR and MLR. Although the kNN approach also performed well with the topological structure of PC-PTFs (i.e., using the logarithm of matric head as extra input variable), it is a non-parametric regression technique with a limited capacity to provide a continuous SWRC prediction.

The evaluation results confirm the superiority of the ANN and kNN approaches in modeling the relationship between soil and water as a complex system even when a limited dataset is available. These findings are significantly important for tropical delta regions, where only very few limited data are available for PTFs development, despite the growing demand to develop soil databases in such regions. Due to the black-box and user-defined natures of ANN techniques, the practical implementation of this technique has not usually been transparent to all PTF users. The usage of the kNN method, on the other hand, would have greater benefits because of its flexibility, simplicity, accuracy and capacity to append new observations in training data sets without the need to redevelop the models again.

In order to cope with the limitation of PTFs' transferability caused by temporal variation of soil hydraulic properties, incorporating soil properties which reflect the temporal change of soil quality, e.g. soil structure, might be helpful to improve PTF accuracy and reliability. Future research about the improved effects of soil structural information on SWRC estimation in combination with different regression methods is recommended.

Chapter 6

COMBINED EFFECTS OF REGRESSION METHODS AND THE USE OF SOIL STRUCTURE ON PTFS' ACCURACY

This chapter is rewritten based on:

Phuong Minh Nguyen, Jan De Pue, Khoa Van Le, Wim Cornelis (2015). Impact of regression methods on improved effects of soil structure on soil water retention estimates. *Journal of Hydrology* (525), 598-606.

6.1. Introduction

In order to get the information of SWRC in a cost-effective way, indirect estimation of such property from easily measurable or readily available soil data by using PTFs is becoming increasingly popular. Since the indirect estimates are often tainted with considerable uncertainty, many attempts have been devoted to improve the accuracy of these predictive functions. Pachepsky et al. (2013) stated that improvements in PTFs' predictability can be achieved by using more flexible PTF algorithms, adding more significant predictors into PTFs development, and preliminary grouping of soils.

It is widely known that SWRC is a function of soil structure and soil texture (Or and Wraith, 2002). Incorporating soil structure information into texture-based PTFs, therefore, has been reported to improve the accuracy of SWRC-PTFs (Pachepsky et al., 2006). In a recent review on using PTFs for estimating soil hydraulic parameters, Vereecken et al. (2010) concluded that further improvement of PTFs is mainly limited by a lack of new information such as soil structure and questioned how to best include soil structural information into PTFs. Depending on how soil structure characteristics are assessed (e.g., visual morphology description or quantitative soil structural indices), soil structure information can be incorporated directly as PTF's input predictors (Giménez et al., 2001; Lin et al., 1999b; Pachepsky et al., 1998) or as grouping criterion to partition soils into homogeneous subgroups with similar SWRC for specific PTF development (Danalatos et al., 1994; Pachepsky and Rawls, 2003; Pachepsky et al., 2006; Rawls and Pachepsky, 2002; Williams et al., 1992).

In the literature, most of the studies about the effect of soil structure on the accuracy of SWRC-PTFs were carried out via statistical regression technique. With the increasing popularity of data-mining techniques nowadays as a powerful tool for PTF development, it is very useful to examine the impact of various regression methods on the PTF accuracy when categorical soil structure information routinely available from soil survey databases is incorporated into PTFs development. Although other indirect properties to express soil structure such as penetration resistance (Pachepsky et al., 1998), soil aggregation or those extracted from recent developments, like x-ray tomography, spectral induced polarization, and nuclear magnetic resonance (Vereecken et al., 2010), might potentially improve PTF predictions, they are simply not available in most databases.

Support Vector Machines (SVM) and k-Nearest Neighbors (kNN) are two promising approaches employed in PTF research nowadays, due to their flexibility and accurate performance (**Chapter 5**). The kNN approach was applied successfully to develop SWRC-PTFs for soils in both temperate (Nemes et al., 2006a) and tropical regions (Botula et al., 2013). SVM is now considered as a promising alternative to artificial neural networks (ANN) as it can eliminate the local minimum issue, which is the main weakness of the ANN approach. Lamorski et al. (2008) applied SVM to predict soil water contents at various pressure heads using simple basic properties of Polish soils. A similar SVM approach was employed by Twarakavi et al. (2009) to develop parameter-based PTFs for van Genuchten-Mualem models.

It is widely known that the lack of well-defined and extensive data of soil hydraulic properties in the tropics is one of the major limitations dragging the development of SWRC-PTFs behind (Hodnett and Tomasella, 2002). Nowadays, many attempts have been devoted to study soil-water relationships of tropical soils through developing SWRC-PTFs. These PTFs have been derived using rather limited data which represented specific soils in tropical regions; for instance, highly weathered soils on stable landforms (Botula, 2013), recently developed alluvial soils in a dynamic river basin (Chapter 5), and black clayey soils with strong shrinking and swelling characteristics (Patil et al., 2013).

Based upon the above, it is clear that there is a need to explore the interactions of different strategies on the improvement of PTF's accuracy when only limited data sets are available for PTF development. Since each regression approach has its own situation to work best (Hastie et al., 2009), the objective of this study was to investigate whether incorporating categorical soil structure information will improve the accuracy of PTFs developed based on SVM and kNN approaches. This study contributes in addressing the question on how to best include soil structural information into PTFs by testing different combinations of flexible regression approaches with considering categorical soil structural information. It would in turn be valuable to those interested in developing new PTFs with rather limited data sets at hand.

6.2. Materials and methods

6.2.1. Soil data set

A data set of 160 samples which was used to derive the MLR-PTFs in **Chapter 4** was utilized in this study. Information about soil survey descriptions, samplings, and analysis was described in **Chapter 2**.

In the present study, the morphological soil structure description in terms of types of soil structure according to the FAO Guidelines for Soil Description (FAO, 2006) was used as grouping criterion to partition data into three uniform subsets of massive, structured and structureless soils with sample size of 91, 46 and 23, respectively (as applied with MLR in **Chapter 4**). The basic soil properties (i.e., sand, silt, clay content, OC, and BD) were utilized as predictors for SWRC estimation with both SVM and kNN techniques. Descriptive statistics of predictor variables and SWRC of the whole data set and the three structural subsets are presented in **Table 6-1**. The variation of soil texture in the data set is graphically displayed in **Figure 6-1** where different markers represent the samples of different structural subgroups.

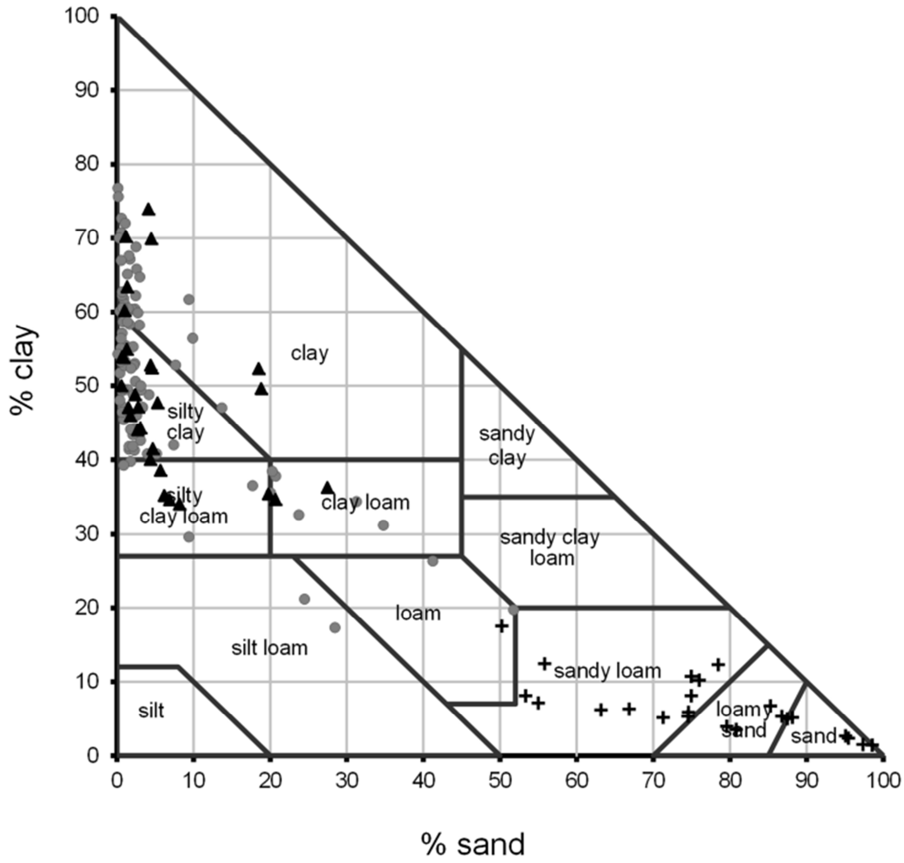

Figure 6-1. Variation of soil texture of the data set (N=160) in the USDA textural triangle, where the circles, triangles and crosses represent soils belonging to massive, structured and structureless subgroups, respectively.

Table 6-1. Descriptive statistics of soil properties in the whole training data set (N = 160 samples), and the specific subgroups of massive, structured and structureless soils (N = 91, 46, 23, respectively).

Soil properties	Whole data set (N=160)				Massive soils (N=91)				Structured soils (N=46)				Structureless soils (N=23)			
	Min.	Max.	Mean	Std.	Min.	Max.	Mean	Std.	Min.	Max.	Mean	Std.	Min.	Max.	Mean	Std.
Organic carbon (%)	0.1	12.3	2.4	2.4	0.2	12.3	2.6	2.5	0.5	11	2.9	2.5	0.08	1.1	0.52	0.3
Bulk density (Mg m^{-3})	0.7	1.9	1.2	0.2	0.7	1.6	1.2	0.2	0.7	1.6	1.2	0.2	1.2	1.9	1.5	0.18
Sand content (%)	0.1	99	16	27	0.1	52	5.4	9.6	0.1	28	5.1	6.1	50	99	77	14
Silt content (%)	0	65	40	14	23	61	43	9.3	22	65	46	9.9	0	39	17	12
Clay content (%)	1.4	77	44	19	17	77	51	12.5	32	74	49	10.3	1.4	18	6.6	4
θ (m^3 m^{-3}) at:																
-1 kPa	0.24	0.74	0.50	0.10	0.34	0.74	0.53	0.08	0.39	0.72	0.52	0.07	0.24	0.41	0.35	0.05
-3 kPa	0.17	0.73	0.49	0.10	0.33	0.73	0.52	0.08	0.39	0.69	0.51	0.07	0.17	0.41	0.32	0.06
-6 kPa	0.12	0.72	0.47	0.12	0.32	0.72	0.51	0.08	0.39	0.65	0.49	0.06	0.12	0.39	0.26	0.08
-10 kPa	0.06	0.71	0.45	0.12	0.30	0.71	0.49	0.08	0.37	0.61	0.47	0.06	0.06	0.34	0.22	0.09
-20 kPa	0.03	0.70	0.41	0.12	0.29	0.70	0.45	0.08	0.35	0.56	0.43	0.05	0.03	0.27	0.17	0.08
-33 kPa	0.03	0.67	0.37	0.12	0.27	0.67	0.42	0.08	0.31	0.53	0.41	0.05	0.03	0.24	0.14	0.06
-100 kPa	0.03	0.58	0.32	0.11	0.22	0.58	0.35	0.07	0.23	0.48	0.36	0.06	0.03	0.17	0.09	0.04
-1500 kPa	0.02	0.43	0.24	0.09	0.11	0.43	0.27	0.05	0.19	0.38	0.27	0.05	0.02	0.09	0.06	0.02

θ is volumetric water content (m^3 m^{-3}) retained at different matric potentials. Min., Max, Std. are the minimum value, maximum value, and standard deviation of soil variables.

6.2.2. Methods to build PTFs

6.2.2.1. Support Vector Machines

In this study, the SVM implementation to derive point PTFs for SWRC estimation was done by R statistical language (R Core Team, 2014). The most commonly used kernel, i.e., the radial basis kernel, which was applied in the works of Lamorski et al. (2008) and Twarakavi et al. (2009), was selected to build our SVM-models. The optimal hyper-parameters of SVM models (i.e., regularization parameter, C; regression precision, ε; and kernel parameter, γ) were estimated using a thorough grid-based search approach (Hastie et al., 2009). A 10-fold cross-validation was applied to find these optimal meta-parameters, i.e., during the search, various sets of C, γ, ε were tried and the one with best cross-validation accuracy was picked and used for SVM-PTFs calibration.

6.2.2.2. k-Nearest Neighbors (kNN)

The basic idea of the kNN technique, named similarity-based technique by Nemes et al. (2006a), is finding the *k* number of nearest neighbors from a reference dataset for each soil in the test dataset in terms of 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 and test dataset following a specific procedure (Nemes et al., 2006a). As soil properties can differ in their order of magnitude or range, rescaling is done to ensure that different input attributes will receive equal weight. Soils in the reference data set 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 particular procedure. Once the nearest neighbors 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 neighbors.

For kNN estimation, two design parameters of the *k*-NN algorithm were defined and used for the estimation procedure, namely the *k* and *p* terms. In this study, we use the proposed formula of Nemes et al. (2006a) for determining them.

More methodological and calculation details on the whole procedure can be found in the works of Botula et al. (2013), Nemes et al. (2006a, 2006b). The kNN algorithm used in this study was adapted from the variants developed by Nemes et al. (2006a) and Botula et al. (2013). The implementation of the kNN algorithm was done in the MATLAB R2014a environment (the MathWorks).

6.2.3. Evaluation criteria

The evaluation of PTF accuracy is commonly made and reported through the comparison of PTF estimated and observed values. Three statistical indices were selected to assess the accuracy of the derived PTFs. They are (1) mean of prediction error (ME), a measure of the prediction bias which indicates the over- or under-estimations of the specific model; (2) the root mean square of prediction error (RMSE), a measure of the overall prediction error; and (3) the coefficient of determination (R^2), which indicates the amount of variation in the data explained by the predictive models. The formula to calculate these statistical indices could be referred in section Materials and Methods of **Chapter 5**

For the SVM technique, the predictive algorithms with optimal meta-parameters were calibrated based on the whole data set (N=160) and specific structural subsets (N = 91, 46, and 23 for massive, structured and structureless soil groups, respectively) to obtain the best-fitted SVM models. These models were then applied again to the training data to evaluate their accuracy. In case of the kNN algorithm, which is an instance-based regression method which needs separate test data for evaluation, the accuracy of the kNN models was evaluated based on the leave-one-out (LOO) procedure. As the name suggests, the LOO method is performed by leaving one sample out of the reference/training data set and the remaining samples are used as the reference data to derive the estimation for the leave-out sample (Mucherino et al., 2009). Therefore, N = 159 for the whole data set, and N = 90, 45 and 22 for the massive, structured and structureless subgroups, respectively. The statistical indices defined as measures of PTFs accuracy were calculated based on all observations in the training data sets.

6.2.4. Statistical analysis

Moreover, as our interest lies in the efficiency of incorporating categorical soil structure information on the accuracy of PTFs developed by different regression techniques (i.e., kNN and SVM), the mean square error of the two factors (i.e., with and without incorporating soil structure in PTFs development) of the particular regression approach was assessed by a pairwise comparison test based on the t distribution (LSD test).

For testing the difference of soil properties between specific structural subgroups, one-way Anova (for 3 groups comparison) or independent t-test (for 2 groups comparison) at 0.05 significance level was displayed after the assumption about the normal distribution of soil variables was checked and fulfilled. Otherwise, non-parametric tests were utilized.

6.3. Results

The performance and accuracy of the derived PTFs, in terms of ME, RMSE, and R^2 based on the whole dataset and specific structural subsets is presented in **Table 6-2** and **Figure 6-2**, **Figure 6-3**.

Table 6-2. Mean error (ME) and coefficient of determination (R^2) of point PTFs developed by Support Vector Machine (SVM), and k-Nearest Neighbors (kNN) techniques based on the whole data set (before grouping) and three structural subsets (after grouping).

Methods to derived PTFs	SVM		kNN	
	before grouping	after grouping	before grouping	after grouping
	ME (m^3 m^{-3})			
PTF at: -1 kPa	-0.0004	-0.0008	-0.0002	-0.0013
-3 kPa	0.0006	0.0003	-0.0007	-0.0020
-6 kPa	-0.0013	0.0006	-0.0016	-0.0038
-10 kPa	0.0029	-0.0029	-0.0018	-0.0042
-20 kPa	-0.0009	-0.0060	-0.0024	-0.0045
-33 kPa	-0.0030	-0.0071	-0.0026	-0.0045
-100 kPa	-0.0060	0.0000	-0.0021	-0.0035
-1500 kPa	-0.0004	-0.0006	-0.0005	-0.0010
	R^2			
PTF at: -1 kPa	0.79	0.74	0.68	0.66
-3 kPa	0.78	0.78	0.71	0.71
-6 kPa	0.79	0.85	0.76	0.77
-10 kPa	0.82	0.88	0.79	0.80
-20 kPa	0.83	0.87	0.79	0.79
-33 kPa	0.81	0.83	0.76	0.75
-100 kPa	0.82	0.79	0.73	0.71
-1500 kPa	0.81	0.81	0.76	0.75

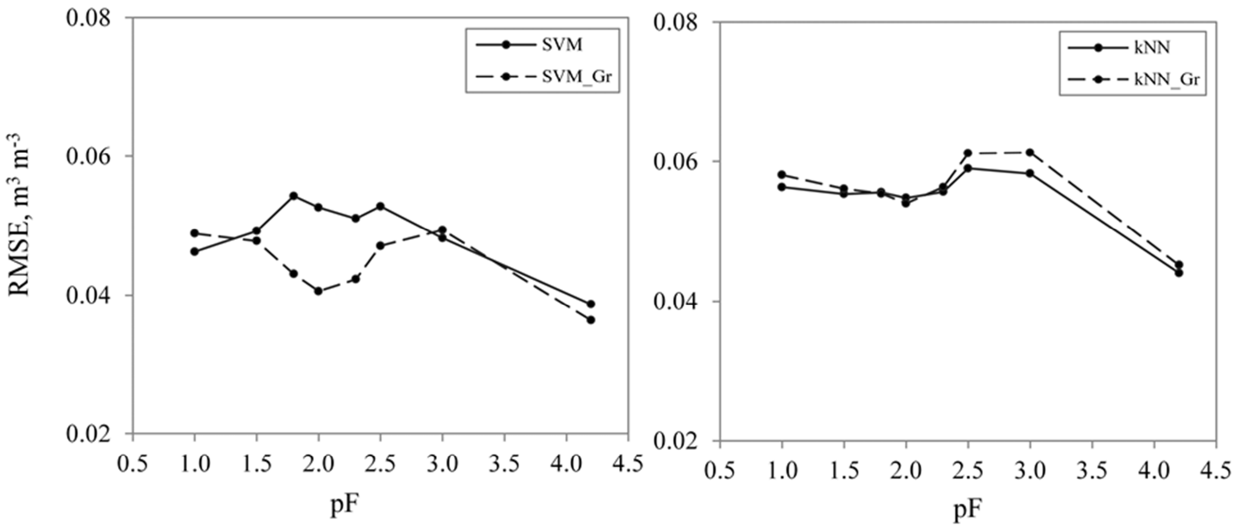

Figure 6-2. Variation of RMSE as a function of pF (pF = $\log_{10}(h)$, where h is the potential head expressed in cm of water) before and after grouping by soil structure for PTFs developed by Support Vector Machines (SVM) and k-Nearest Neighbor (kNN) approaches.

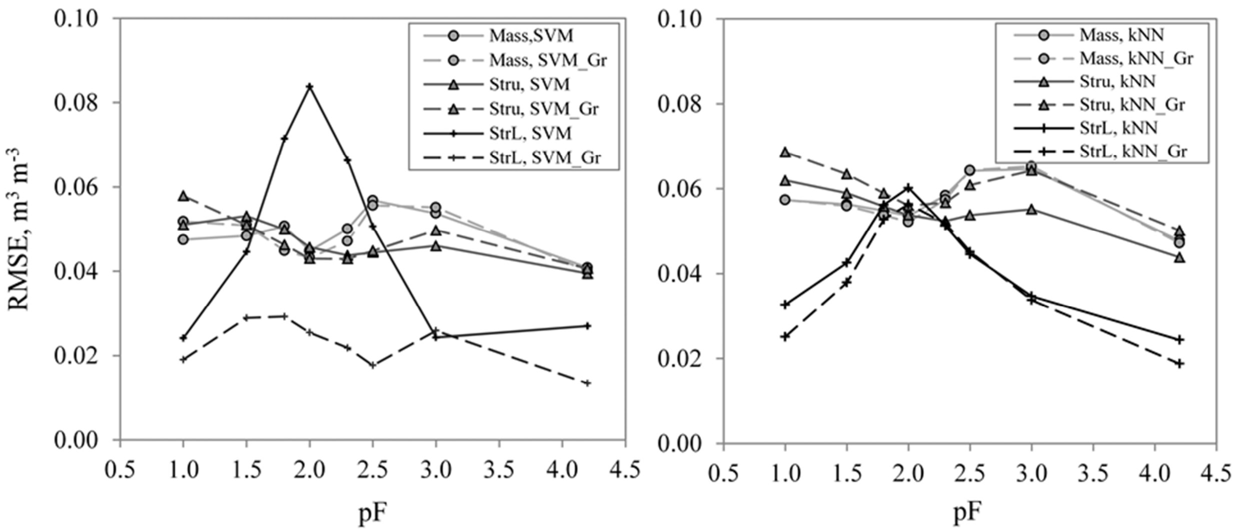

Figure 6-3. Variation of RMSE as a function of pF (pF = $\log_{10}(h)$, where h is the potential head expressed in cm of water) of specific subgroups ('Mass' is the massive soils group, 'Stru' is the structured soils group, and 'StrL' is the structureless soils group). SVM, kNN, SVM_Gr, kNN_Gr are the PTFs derived by Support Vector Machines and k-Nearest Neighbor approaches based on the whole data set and specific structural subsets, respectively.

6.3.1. Effect of soil structural grouping on SVM-PTFs

The results show that incorporating soil structural information into PTF development does improve the accuracy of derived SVM-PTFs in the range of matric potentials between -6 kPa (pF1.8) and -33 kPa (pF2.5). The coefficient of determination (R^2) of the point PTFs calibrated by using structure-based subgroups was enhanced, e.g., the R^2 values in this range vary from 0.79 to 0.83 before grouping and increase to 0.83 to 0.88 after grouping by soil structure (**Table 6-2**). Accordingly, the RMSE values of PTFs derived from soil structure-based subsets at those matric potentials are significantly lower ($p \leq 0.05$) than those of PTFs derived based on the whole data set (RMSE equal 0.051 – 0.053 $m^3\ m^{-3}$ before grouping, and 0.043 – 0.047 $m^3\ m^{-3}$ after grouping by soil structure) (left panel of Fig. 2). The RMSE of SVM PTFs at other matric potentials did not expose any significant difference before and after grouping by soil structure.

When considering RMSE per subgroup, the left panel of **Figure 6-3** displays a substantial improvement for structureless soils, with values being significantly different at all matric potentials except at -1 and -100 kPa (pF1.0 and pF3.0, respectively). Differences in RMSE before and after grouping were not significant for the massive and structured soil subgroups.

In concert with the substantially small values of mean prediction error (ME ≈ 0) (**Table 6-2**), the scatter plots of the SVM-PTFs calibrated based on the data sets before and after grouping (**Figure 6-4**) display a good agreement between the measured and PTF predicted soil water content. Most of the points in both graphs are scattered around the 1:1 reference line and do not exhibit much bias. Moreover, as manifested previously with RMSE values, the scatter points in the plot after grouping by soil structure (right panel of **Figure 6-4**) get closer to the 1:1 reference line, particularly at smaller water contents. This visual aspect is well concordant with the significant improvement in the structureless soils group whose soil water content at any matric potentials is smaller than in the massive and structured soils.

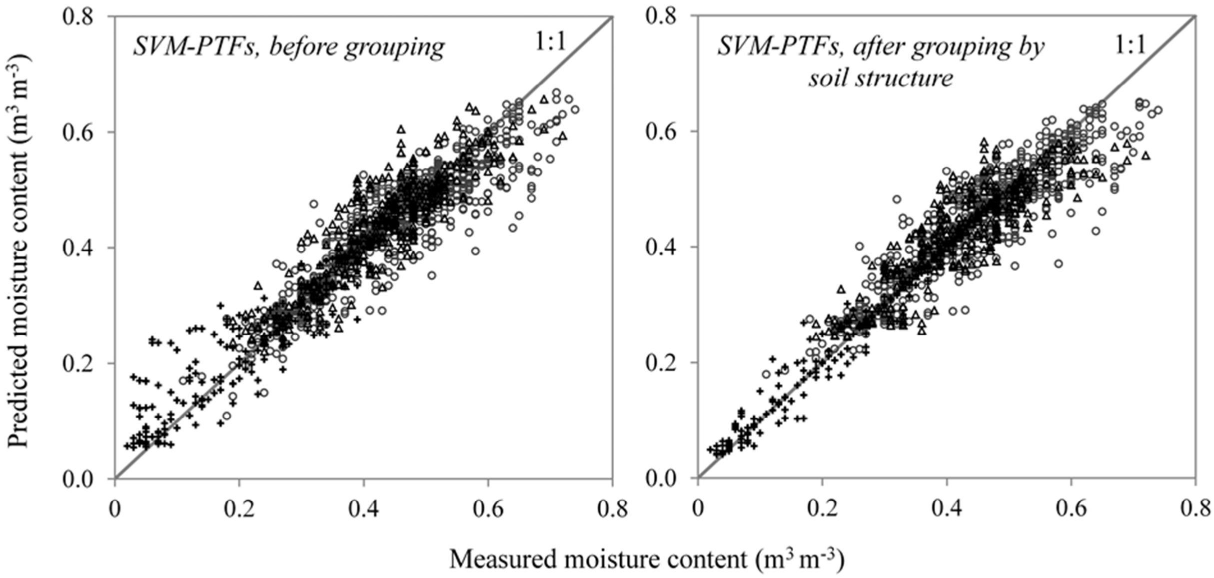

Figure 6-4. Scatter plots of measured *vs.* SVM-PTFs' predicted soil water content before and after grouping by categorical soil structure. The circles, triangles and crosses represent soils belonging to massive, structured and structureless subgroups, respectively.

6.3.2. Effect of soil structural grouping on kNN-PTFs

Contrary to the results of the SVM approach, all error measures with the kNN models show that no improvement in PTF predictability was obtained when soil structure was incorporated as grouping criterion for specific PTF development. R^2 values of kNN models developed using the whole data set were comparable and in some cases even higher than those derived from homogeneous structure-based subgroups (**Table 6-2**). The R^2 values are in the range of 0.68 - 0.79 before grouping and of 0.66 - 0.80 after grouping by soil structure. Similar variation was obtained with RMSE (right panel of **Figure 6-2**) with no significant differences observed before (RMSE varies between 0.044 and 0.059 m^3 m^{-3}) and after grouping by soil structure (RMSE ranges from 0.044 to 0.059 m^3 m^{-3}); p values of the paired t-test were larger than 0.05. The graphical presentation of the correspondence between observed and PTF predicted soil water content displays more or less a similar pattern before and after grouping by soil structure (**Figure 6-5**).

Detailed investigation about the contribution of specific-structured subgroups on average RMSE (right panel of Fig. 3) shows similar variation of the prediction errors of PTFs derived based on the information of the whole data set and specific-structural subsets.

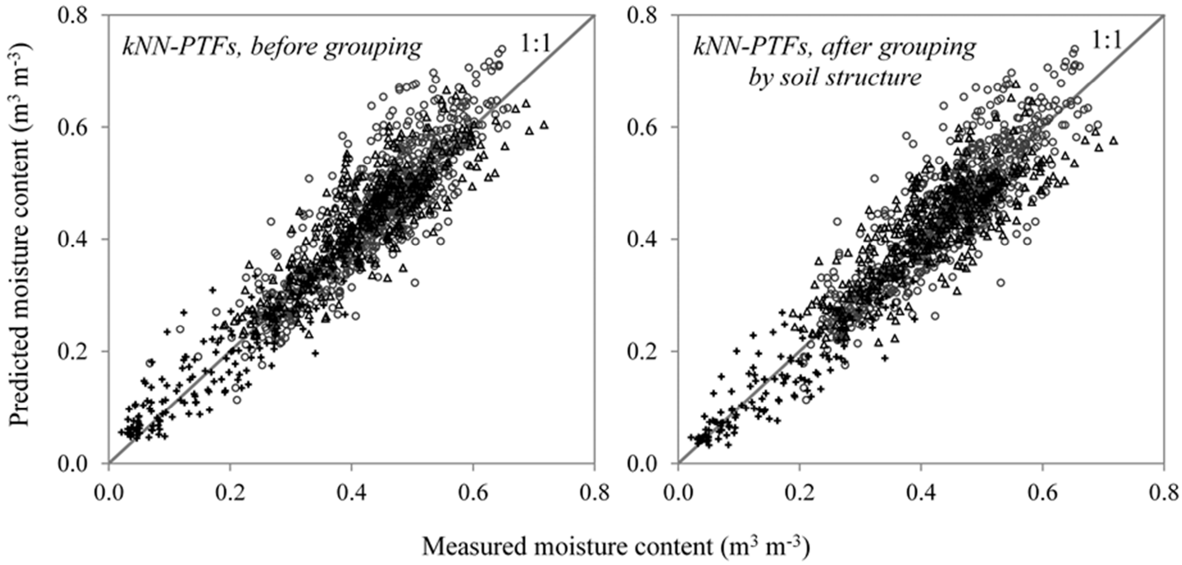

Figure 6-5. Scatter plots of measured soil moisture content vs. kNN-PTFs' predicted soil moisture content before and after grouping by categorical soil structure. The circles, triangles and crosses represent soils belonging to massive, structured and structureless subgroups, respectively.

6.4. Discussions

The improved accuracy of SVM models calibrated by structural subgroups in the present study is in good accordance with the results obtained previously with MLR (**Chapter 4** and Williams et al., 1983), and regression trees (Rawls and Pachepsky, 2002). Using categorical soil structure information as grouping criterion resulted in more homogeneous subgroups with less variation in soil properties (e.g., soil texture and SWRC) (**Table 6-1** and **Figure 6-1**). The SVM PTFs trained on specific structural subsets were more accurate in predicting SWRC in the intermediate range (left panel of **Figure 6-2**). Such improvements are primarily attributed to the good performance of PTFs of structureless soils (left panel of **Figure 6-3**). As logically perceived, the SWRC of the structureless group and its correlation trend to basic soil properties, were significantly different from those of the massive and structured subgroups, particularly at intermediate range. This is illustrated in **Figure 6-6**, where soil water content at air-entry pressure (i.e., pF1), field capacity (FC) (i.e., pF2.5), and permanent wilting point (PWP) (i.e., pF4.2) are plotted as a function of clay content. The linear SVM regression plane calibrated on the whole data set which is biased to fined-textured soils (**Figure 6-1**) seems to imprecisely describe the relationship between the predictor variables and SWRC of structureless soils. General SVM-PTFs, therefore, provide less accurate prediction of SWRC at intermediate range of structureless soils compared to the PTFs calibrated by that specific subgroup.

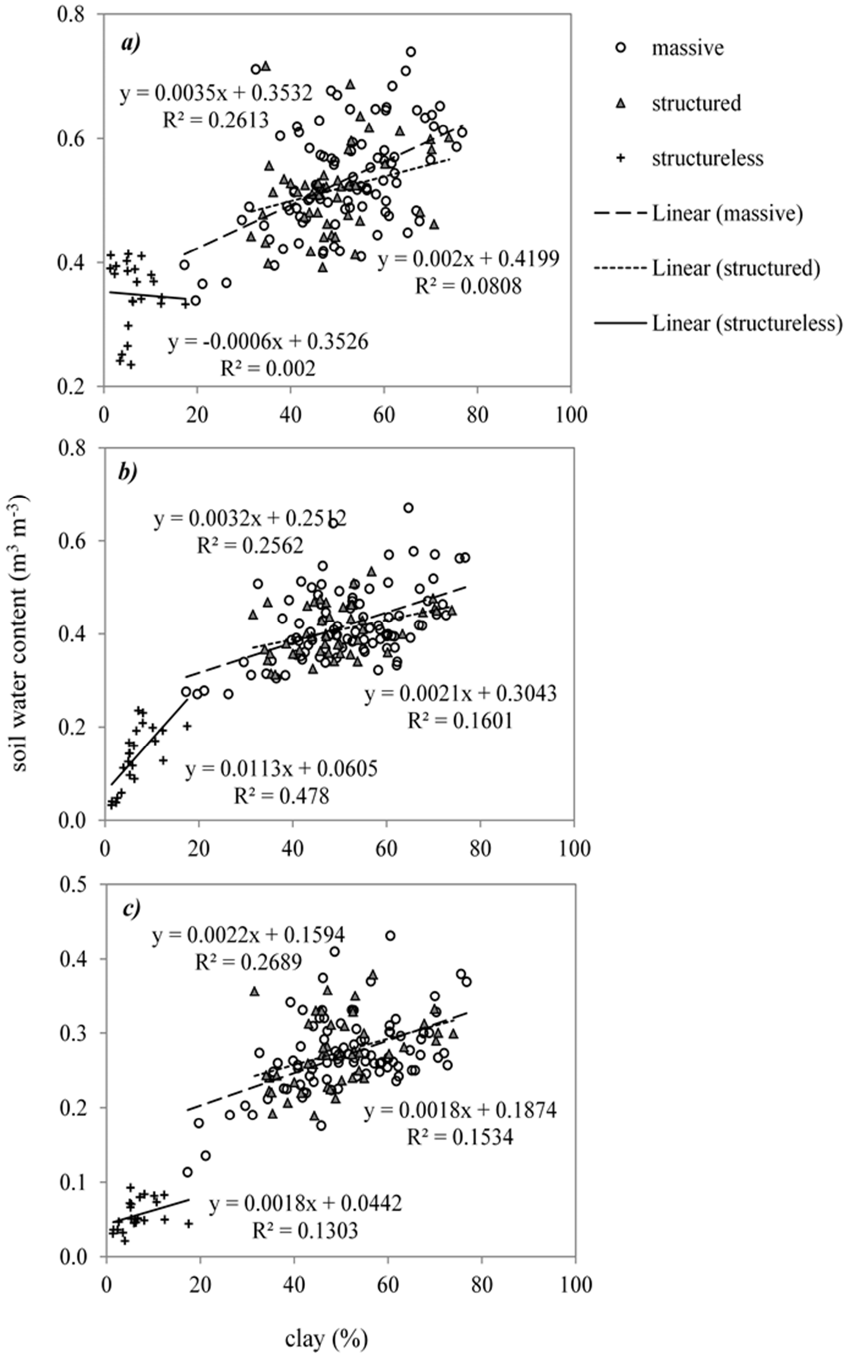

Figure 6-6. Variation of soil water content (m^3 m^{-3}) at (a) air entry (pF 1), (b) field capacity (pF 2.5), and (c) permanent wilting point (pF 4.2) as a function of clay content, where the circles, triangles and crosses represent soils belonging to massive, structured, and structureless subgroups, respectively.

Moreover, it is widely known that soil water content in the capillary zone of SWRC (i.e., intermediate range) is strongly defined by pore-size distribution, or 'soil structure' (Radcliffe and Šimůnek, 2010). For structureless soils, primarily coarse and medium textured soils, the capillary pore-size distribution can be satisfactorily described by particle size distribution (Arya and Paris, 1981). Hence, SVM PTFs calibrated on structureless data provide better prediction of SWRC at these matric potentials.

In the air-entry range (e.g., at -1 kPa) of the soils, the lack of improvement might demonstrate that the variation in soil water content is mainly explained by total porosity (determined by BD and OC), rather than by pore-size distribution which is reflected by the higher Pearson correlation between soil water content at -1 kPa and bulk density (r = 0.72) as compared to that at matric potentials below -6 kPa (r < 0.67). **Figure 6-7**, where soil water content at pF1 is plotted as a function of bulk density, visually supports the strong correlation between bulk density and water content at air-entry pressure. All three structural soil groups have similar slope of regression function with BD, hence grouping does not play a role at that point.

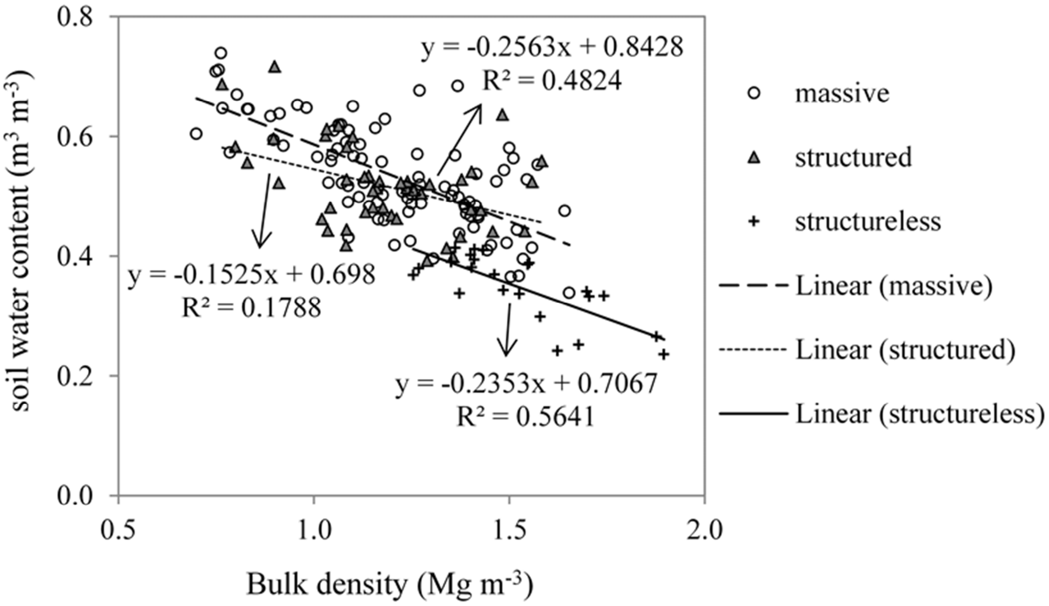

Figure 6-7. Variation of soil water content (m^3 m^{-3}) at pF 1 as a function of bulk density, where the circles, triangles and crosses represent soils belonging to massive, structured and structureless subgroups, respectively.

In case of massive and structured soil groups, an improved accuracy of PTFs calibrated by specific structure types is not clearly observed in both the SVM and kNN approaches. Indeed, exploratory data analysis did not display a significant difference in SWRC between these two groups ($p > 0.05$). The soils in these two groups cover a wide range of soil hydro-physical properties which overlap each other.

It is important to note some specific features of soil structural properties in the tropical delta where the study was conducted in order to explain why the hydro-physical behaviors of massive and structured soils are not so distinguishable. First of all, due to the recent origin of the soils formed in the river delta basin, their majority shows weak structural development (Le, 2003). In the used data set, the soils with certain degree of structural development (i.e., structured group) predominantly consist of weak-developed granular aggregates. Such type of structural unit is, to some extent, close to the massive structure of the surface horizons (i.e.., those with high content of OC and low BD) in terms of pore-size distribution. Moreover, as it is widely recognized, structural development of such young delta soils is mainly promoted by agricultural activities, hence the samples taken from paddy rice fields (the dominant agricultural practice in the region) show a massive structure, while soils under upland crops exposed some extent of structural development although in a weak grade. Since the land under upland crop cultivation is limited in the Mekong Delta, samples taken from such fields are underrepresented in the data set as compared to samples taken from paddy fields.

On the other hand, typical soil preparation for lowland paddy rice cultivation, i.e., ploughing and puddling under submerged condition, results in a puddled layer in the surface and a compacted plow-layer underneath (Sharma and De Datta, 1986). Although both horizons display massive structure type, SWRCs between the two horizons are different, particularly in the wet range. Actually, within the massive group in the present data set, the surface and subsurface horizons are significantly different in OC, BD and SWRC at pF1 and pF1.5 ($p < 0.05$). The massive Ap horizons have significantly higher OC content and lower BD compared to the massive B horizons underneath, hence total soil porosity which is mainly defined by OC and BD is also higher as manifested by the higher water content in the wet range of SWRC (e.g., pF1 and pF1.5) of massive upper soil horizons.

In addition, the performance of the kNN models can be logically explained by the predictive algorithm of non-parametric kNN approach as well as the characteristics of training data as discussed hereinbefore. Unlike other parametric techniques as MLR and SVM, the non-parametric kNN technique does not rely on any mathematical functions to estimate SWRC. Application of the k-NN technique means identifying and retrieving the nearest (most similar) stored objects in the reference data to the target object in terms of selected input variables (i.e., soil texture, BD and OC). As shown in **Figure 6-1** and **Table 6-1**, input variables of kNN-PTFs in the present study (i.e., soil texture, BD, and OC) are inherently clustered in correspondence with soil structure type, particularly structureless soils. The coarse and medium textured soils of structureless group contain significantly low OC content, and high BD. Such inherently separations of input variables make the selection of nearest neighbors from the whole data or structureless soil group are more or less similar. Hence, grouping by soil structure does not expose any improvement in the kNN PTF accuracy for the structureless subgroup as it does with SVM models.

Moreover, the kNN is proven to be a highly data-demanding technique which needs a large soil database to obtain stable and accurate estimations (Hastie et al., 2009), hence grouping by soil structure leads to more homogeneous soil groups of smaller size which in turn might increase the bias of the PTF prediction. Indeed, the kNN models provide the prediction by locally averaging the response of the nearest neighbors around the target soils. The prediction with the kNN method at the boundary of the data space is hence suffers from large bias as the nearest neighbors mostly lie on one side of the target soils. As it is manifested with the bias measure (i.e., ME), kNN estimations based on structure subgroups have the values of ME of two to five times higher than those calibrated by using the whole data set (**Table 6-2**).

Indirect incorporation of soil structure as grouping criterion did, at some extent, expose an improvement in the accuracy of the SVM PTFs, but not in that of the kNN PTFs in the present study. Utilizing semi-quantitative approaches of visual field assessments for soil structural quality (e.g., Visual soil assessment (VSA), Shepherd (2009); Soil Quality Scoring Procedure (SQSP), Ball and Douglas (2003); and Visual Evaluation of Soil Structure (VESS), Ball et al. (2007)) as a direct 'soil structure' predictor might be another strategy for incorporating soil structural information. Recently, Pulido Moncada et al. (2014) found that soil structural index and number of earthworms, two parameters of VSA method of Shepherd (2009), are important predictors of saturated hydraulic conductivity, Ks. These visual soil assessment methods are now gaining popularity in evaluating soil quality with scores being recorded as temporal indicators of soil structural quality change (Pulido Moncada, 2014). Direct incorporation of descriptive soil structure information under scoring systems as continuous predictor of SWRC might improve the predictive capacity of PTFs, as well as mitigate small data set sizes caused by preliminary grouping based on qualitative soil structural information.

6.5. Conclusions

The present study shows that with the availability of rather limited training data set for PTF development, the effect of using categorical soil structure information in SWRC estimation depends on the used regression techniques. We only found an increased PTF accuracy resulting from soil structural information when using the parametric SVM method, but not with the non-parametric kNN technique. These findings are probably related to the predictive algorithms of the two techniques, as well as the characteristics of the training data. Indeed, the parametric regression approach tested here (i.e., SVM) offers an estimation based on the mathematical functions which are empirically derived using the entire training data. By clustering the soils with similar soil structural morphology, preliminary grouping helps to decrease the variability within homogeneous groups. Therefore, more accurate description about the structural dependency between predictors and predictands can be performed by SVM models. In case of the non-parametric kNN technique, the accuracy of the PTFs estimation seems more related to the selected ‘nearest neighbors’ and the size of the reference data set. Moreover, since the variability of kNN input variables (i.e., soil texture, OC, and BD) in the present data set is inherently clustered in correspondence with soil structure types, grouping by soil structure does not play a role in such situation.

Overall, we found that ‘simply’ grouping data in three subsets, even when data are limited, might be an option for improving the accuracy of PTFs. However, at least when datasets are limited in size, we would suggest using a parametric SVM (or related) approach rather than a non-parametric kNN approach when developing grouping-based PTFs. Our findings might need to be confirmed by other studies with other (and larger) data sets. More research on finding thresholds for data set size is warranted, as this is yet unknown.

Utilizing the numerical score provided by visual and tactile soil structural assessments might probably be a promising strategy of PTFs improvement. Further research is needed as well to investigate the ability of these visual scores as direct numerical predictors for increasing the accuracy and reliability of SWRC prediction to PTFs.

Chapter 7

FUNCTIONAL EVALUATION OF PTFS' PERFORMANCE USING AGRO-HYDROLOGICAL MODELS AQUACROP AND HYDRUS-1D

7.1. Introduction

Problems of fresh water shortage and ground water contamination are arising on a global scale (Georgoussis et al., 2009), particularly in areas of intensive rice production due to the issues of over-irrigation and over-amendment of chemical fertilizers and pesticides to maintain high yields. Facing these environmental problems and finding remedies, requires the understanding of several important processes that take place in soil. However, direct field experimentations of soil-water system under different management scenarios are often time-consuming, costly and sometimes can even be risky as they can carry undesired and hazardous risks. Suitable soil-water models can offer alternatives that are quicker and easier to execute and may give at least indicative answers about trends that are expected to occur in reality without any risk to the environment (Nemes et al., 2010).

Last several decades, various agro-hydrological models have been developed and have become indispensable tools for studying, for example, crop growth, soil erosion, catchment hydrology, effects of climate change on soil and water resources, and agricultural production. The models are intended to quantify and integrate the most important physical, chemical, biological, and hydrological processes in the unsaturated zone with the aim of underpinning sustainable natural resources management (Wösten et al., 2013). Depending on the specific research purposes as well as the data availability, different types of soil-water models, ranging from simple crop models which solve the soil-water balance using a bucket approach to complex hydrological models which are based on a numerical solution of Richards' equation for simulating water and solute movement, can be effectively used to study several important soil-water processes. Regardless of the model's complexity, all soil-water models require inputs of soil hydraulic properties (i.e., soil-water retention and hydraulic conductivity characteristics). Parameterization of soil hydraulic inputs for fine-scale modelling can be obtained through sampling and direct measurement, but such procedure is hard to perform for fulfilling the data requirement of large-scale research projects due to the highly temporal and spatial variability of soil properties, as well as the difficulty and high labour costs of measurement methods of particularly soil-water retention curve.

The difficulty in obtaining measurements of soil hydraulic parameters, as well as the increased demand for input data in models, urged scientists to consider alternative data sources. Pedotranfer functions (PTFs), relating easily measurable or readily available soil properties to the needed soil hydraulic parameters, can deliver candidate approximations for the required models' hydraulic information (Christiaens and Feyen, 2001). Although PTFs seems to be the answer to the shortage of required data for practical model application, the use of indirect methods always carries some level of uncertainty. Given the sensitivity of mathematical models to hydraulic parameter's variation, the use of indirectly PTF's estimated values could probably jeopardize the results of model simulation, and consequently, support decisions or planning guidance for sustainable resources management. This matter reasonably requires cautious evaluation of PTFs performance with respect to specific model applications.

Several studies have been carried out so far to address such kind of problem by evaluating the performance of investigated PTFs with respect to their utility in agro-hydrological modeling, e.g., the works of Christiaens and Feyen (2001); Leenhardt (1995); Moeys et al. (2012); Nemes et al. (2010); Vereecken et al. (1992) to mention a few. Espino et al. (1996) used a deterministic model approach to evaluate PTFs performance on the outcomes of the mathematical model SWATRER by comparing observed and simulated values of soil-water content and pressure heads. Their results revealed that major differences can exist among the used PTFs and the authors ended up providing several reasons why PTFs' use should be done with caution and critical eyes. Similarly, Gijsman et al. (2002) evaluated the PTFs developed for soils in US using the CROPCRO-Soybean model. They also observed a worrisome variability among the indirect methods in simulated crop yield, hence cautions about the use of PTFs were also drawn to crop modelers. Follow the same token, Georgoussis et al. (2009) investigated the effectiveness of PTFs' use on simulating irrigation scheduling of soils cultivated sugar beet and cotton using SWBACROS model. They found that the use of PTFs resulted in an over-prediction of the observed soil-water contents, and as a result of that, determination of irrigation event and quantity was significantly altered.

In the same manner to these functional evaluations, the objective of this study was to investigate the effect of replacing laboratory derived soil-water retention parameters by PTFs' estimated parameters on the functional outcomes of agro-hydrological models, i.e., soil-water content and (relative) rice yield, for four different soil types and for climate conditions prevailing in the Vietnamese Mekong Delta. To that end, the best performing regional PTF derived in previous studies, i.e., based on Support Vector Regression with incorporation of soil structural information (in **Chapter 5** and **6**), was selected and evaluated in terms of simulating rice yield and net irrigation requirement using the conceptual crop-water model AquaCrop, as well as time series of soil-water contents and degree of water stress (or relative rice yield) using the physically-based hydrological model Hydrus-1D. Model outcomes were compared with those using laboratory measured data and predicted data from globally presented PTFs (i.e., the ones embedded in the models as alternative in cases of missing data). The investigations were carried out for two rice growing seasons within one year.

7.2. Materials and methods:

7.2.1. Study area and soil data of investigated profiles

The study was conducted in VMD, one of the biggest rice production regions in Vietnam. Due to its favorable climatic conditions and fertile soil resources, intensive agriculture is the major form of land use with particularly paddy rice production being a crucial sector for the socio-economy of the region and the country. From around 1980 onwards, intensive rice cultivation with up to three rice crops per year was introduced and applied widely in the region. This system continues to expand to ensure the food production for both domestic use and export markets (Le, 2003), though recent studies have shown that it leads to a decline in soil quality and in rice productivity (Linh et al., 2015a; Linh et al., 2015b).

The functional investigation of the selected PTFs' performance was carried out based on information of four soil profiles which described and collected in the period of 1997-1999 by the department of Soil Science, Can Tho University. These soil profiles represent the major soil groups which are mainly exploited for agricultural production in the region. Since the soil profile description, land use history, and social-economic aspects of the studied sites have been described in detail in Le (2003), we only recall here some relevant information that was highly associated with the functional evaluation study.

The first selected soil profile was located in Hau Giang province (hereafter called HG profile) where acid sulfate soil is the predominant soil group. The investigated profile has very typical characteristics for acid sulfate soil (Typic Sulfaquept), e.g., high content of organic matter (OC = 7.75%) in the top horizon, low pH throughout the profile (pH_{H2O} < 3.5), and high Al^{3+} content in both soluble and exchangeable forms. The diagnostic horizons in soil profile display different degree of structural development such as massive in Ah surface horizon and weak to moderate coarse prismatic in Bgj and Cj horizon. Originally, such kind of soils were not suitable for agriculture, but after a long period of remediation (i.e., flushing of acidity, adding amendments to increase soil pH and balance major nutrients), the soil has become suitable for rice and other cash crop cultivation. On severe acid sulfate soils, natural vegetation is normally found.

The second profile was located in the coastal area of Soc Trang province (hereafter called ST profile) where soil was affected by saline water from annual tidal movement (Thapto-Hydraquentic Aeric Halaquept), hence, high electrical conductivity (ECe) was noticed throughout the profile. Coarse angular blocky soil clods were observed in Ap horizon, while massive structure was found in underlying Ab horizon. The third horizon (Cgb) displayed a weak, coarse prismatic structure. As the profile was dug in a raised bed of sugarcane and other vegetables, a topsoil was fully disturbed.

The third profile was located in Can Tho province (CT profile) where Hau River, one of the main branches of the Mekong river, is passing through. The soil is classified as a recent developed alluvial soil (Fluvaquentic Endoaquept) with typically a stratified differentiation in soil texture among diagnostic horizons. Massive structure was observed in the first (Ap) and third (Cr) horizons, whereas weakly developed coarse to medium angular blocky structure was found in Bg horizon. Due to the rather fertile soil conditions and sufficient fresh water supply for irrigation, the region has long been exploited for intensive paddy rice and other vegetables cultivation.

The fourth profile was of a highly weathered soil (Aquic Haplustept) in a mountainous area in An Giang province (AG profile) and built up by sandy and colluvial materials. As a result of that, the whole soil profile was structureless and displayed a very shallow effective rooting depth, approximately 20 cm.

The soil physical, chemical, and hydraulic properties of diagnostic horizons of four investigated profiles were determined and presented in **Table 7-1**. All soil properties were analyzed using standard methods, e.g., Dane and Topp (2002); Page (1982) which have been widely used for laboratory soil analysis. Specifically, soil bulk density (BD) was measured on 100 cm^3 undisturbed soil cores by the gravitational method of Grossman and Reinsch (2002), soil organic carbon by the wet combustion method (Walkley and Black, 1934), particle size distribution by the sieve-pipette method (Gee and Bauder, 1986), and the soil water retention curve (SWRC) by standard desorption techniques with hanging water column and pressure plates (Hillel, 1980). Specifically, water contents were determined gravimetrically on 100 cm^3 undisturbed soil cores subjected to matric potentials of 1, –3, –6, –10, –20, –34, –100 kPa, and on a disturbed soil sample at –1500 kPa. The widely-used van Genuchten equation (1980) (Eq. 7-1) was fitted to the eight observed water retention data pairs, with *m* taken as 1-1/*n*.

$$\theta(\psi) = \theta_r + (\theta_s - \theta_r)\left(\frac{1}{1+(\alpha|\psi|)^n}\right)^{\left(1-\frac{1}{n}\right)} \tag{7-1}$$

7.2.2. Simulation models

In this study, the evaluation of the PTFs' utility was performed with a conceptual crop-water model (AquaCrop; Steduto et al., 2009) and a physically-based hydrological model (Hydrus-1D; Šimůnek et al. (2008)). Both models are widely used in a context of agricultural water management and rice production, depending on the study's objectives (see below). Assessment of model specification in capturing the reality of important soil processes is beyond the objectives of this study, since no field observations on model outcomes were present and our main objective is to evaluate the response of agro-hydrological models to changes in soil hydraulic inputs stemming from different approaches, i.e., laboratory determined water retention curves and different PTFs. Brief introductions of the two agro-hydrological models used to evaluate the functional performance of the PTFs are represented below.

7.2.2.1. AquaCrop model

FAO's crop-water model AquaCrop is a canopy-level and engineering type of agro-hydrological models mainly focused on simulating the attainable crop biomass and harvestable yield of major herbaceous crops in response to water availability (Steduto et al., 2009). The model achieved significant improvement in accuracy over the approach of Doorenbos and Kassam (1979) (Eq. 7-2) while maintaining its adequate simplicity and robustness.

$$\left(1-\frac{Y_a}{Y_x}\right)=K_y\left(1-\frac{ET_a}{ET_x}\right) \tag{7-2}$$

where Y_x and Y_a are the maximum and actual yield, ET_x and ET_a are the maximum/potential and actual evapotranspiration, and Ky is crop yield response factor.

As clearly described by Steduto et al. (2009), AquaCrop model evolves from the previous FAO water production function (Eq. 7-2) by:

(i) dividing evapotranspiration (ET) into soil evaporation (E) and crop transpiration (T), to avoid the effect of the non-productive consumptive use of water (E).

(ii) obtaining biomass (B) from the product of water productivity (WP) and cumulated crop transpiration which is actually the core of the AquaCrop growth engine:

$$B = WP * \sum T \tag{7-3}$$

(iii) expressing the final yield (Y) as the product of B and harvest index (HI)

$$Y = HI * B \tag{7-4}$$

Table 7-1. Characterization of soil physical, chemical and hydraulic properties of four studied profiles.

Profile	Horizon	depth (mm)	Sand (%)	Silt (%)	Clay (%)	BD (Mg m^{-3})	OC (%)	FC (m^3 m^{-3})	PWP (m^3 m^{-3})	AWC (m^3 m^{-3})	DW (m^3 m^{-3})	van Genuchten parameters of SWRC				
												θ_r	θ_s	α	n	m
	Ah	0-45	1	34	65	0.86	7.75	0.49	0.25	0.24	0.12	0.00	0.61	0.008	1.21	0.17
HG	Bgj	45-90	2	34	64	1.03	0.8	0.46	0.21	0.25	0.15	0.00	0.61	0.009	1.22	0.18
	Cj	90-120	2	37	61	0.89	0.73	0.48	0.23	0.25	0.16	0.12	0.64	0.007	1.33	0.25
	Ap	0-45	3	42	55	1.33	0.8	0.45	0.21	0.24	0.03	0.00	0.48	0.001	1.31	0.24
ST	B	45-95	1	45	54	1.07	1.02	0.50	0.22	0.28	0.10	0.00	0.60	0.005	1.22	0.18
	Cg	95-120	1	47	52	1.03	1.05	0.51	0.22	0.29	0.08	0.00	0.59	0.003	1.26	0.21
	Ap	0-22	1	38	61	1.14	3.35	0.43	0.19	0.24	0.12	0.00	0.55	0.006	1.23	0.19
CT	Bg	22-60	1	44	55	1.1	2.14	0.45	0.21	0.24	0.13	0.07	0.58	0.007	1.28	0.22
	Cr	60-120	2	39	59	0.83	2.14	0.49	0.22	0.27	0.18	0.00	0.67	0.010	1.22	0.18
	Ap	0-18	80	16	4	1.65	0.55	0.15	0.04	0.11	0.17	0.00	0.32	0.013	1.50	0.33
AG	Bg1	18-83	79	18	3	1.63	0.03	0.14	0.05	0.09	0.23	0.03	0.37	0.030	1.50	0.33
	Bg2	83-100	71	19	10	1.81	0.22	0.17	0.05	0.12	0.15	0.00	0.32	0.023	1.30	0.23

BD is the soil bulk density (Mg m^{-3});

OC is the soil organic carbon content (%), soil organic matter (OM, %) = OC * 1.724;

FC is the soil water content (m^3 m^{-3}) at field capacity (taken at matric potential of -33 kPa for fined-textured soils, and -10 kPa for coarse-textured soils;

PWP is the soil water content (m^3 m^{-3}) at permanent wilting point (taken at matric potential of -1500 kPa);

AWC is the total available water content (m^3 m^{-3}), AWC = FC – PWP;

DW is the drainable water content (m^3 m^{-3}), DW = SAT – FC, with SAT is the soil water content (m^3 m^{-3}) at saturation and equal to the fitted vG parameters θ_s;

θ_r, θ_s, α, n, m are the parameters obtained by fitting the van Genuchten equation to eight measured points of the soil-water retention curve and used to describe a continuous curve of soil water retention characteristic.

Through the crop cycle, the crop response to water availability was determined through four stress coefficients (K_{ws}), e.g. reduction of canopy expansion rate, closure of stomata, acceleration of canopy senescence, and changes in harvest index. The water stress coefficient, K_{ws}, is a function of water content in the root zone, expressed as a fractional depletion (*p*) of total available water content (*TAW* (mm) is the volume of water the soil can hold between FC and PWP), and its values span a range corresponding to the upper and lower threshold in soil water content for specific crop.

$$TAW(mm) = 100 * (\theta_{FC} - \theta_{PWP}) * Z_r \qquad (7\text{-}5)$$

where θ_{FC} ($cm^3\ cm^{-3}$) and θ_{PWP} ($cm^3\ cm^{-3}$) are soil water contents at field capacity and permanent wilting point, respectively, and *Zr* (m) is depth of the root zone.

Generally, the structure of the model was designed as to include integrated sub-model components such as: the soil, with its water balance; the crop, with its development, growth and yield processes; the atmosphere, with its thermal regime, rainfall, evaporative demand and carbon dioxide concentration (CO_2); and field management, with its major agronomic practices such as irrigation and fertilization. In AquaCrop, the soil is configured as horizons of variable depth and different texture, and is considered as a water storage reservoir whose capacity is defined by the soil water content at field capacity (FC) and permanent wilting point (PWP). For simulating the soil water balance, AquaCrop simulates the change of soil water content by keeping track of incoming and outgoing water fluxes at its boundaries using a simple bucket approach. The model performs a daily water balance that includes the processes of infiltration, runoff, internal drainage within the root zone, root extraction in different depth layers, deep percolation, evaporation, transpiration, and capillary rise.

Similarly to other crop models, AquaCrop can be either used at a fine scale, e.g. as a research tool and for precision agriculture, or at a larger scale, such as for regional food security forecast or predicting potential effects of climate change on crop production in part of a country. Although it was only recently developed, the model has been applied widely to simulate yield response of different crops (e.g. Geerts et al. (2009a) for quinoa; Heng et al. (2009) and Hsiao et al. (2009) for maize;

Farahani et al. (2009) and García-Vila et al. (2009) for cotton; Araya et al. (2010) for teff; Jin et al. (2014) and Toumi et al. (2016) for winter wheat; Katambara et al. (2013) and Mondal et al. (2015) for rice) which are grown under different conditions (salinity stress, full irrigation or water deficit) and for establishing irrigation schedules (Geerts et al., 2009b). Recently, AquaCrop has been used to study the impact of climate change on the yield of rainfed- and irrigated-rice grown in the lower Mekong Basin (Mainuddin et al., 2013), in India (Bhattacharya and Panda, 2013), in Myanmar (Shrestha et al., 2014), and in Central of Vietnam (Shrestha et al., 2016).

7.2.2.2. Hydrus-1D

Hydrus-1D is a one-dimensional, physically-based hydrological model which is widely used for simulating water flow and solute transport in variably saturated/unsaturated soils. Application involves a broad range of steady-state or transient water flows, solute transports, and heat transfer problems (Šimůnek et al., 2012). The Hydrus-1D model can be used for both direct problems when the initial and boundary conditions for all involved processes and corresponding model parameters are known, as well as inverse problems when some of the parameters need to be calibrated or estimated from observed data. The approach to model calibration and validation may vary widely depending upon the complexity of the application.

The HYDRUS program numerically solves the Richards equation (Richards, 1931) (Eq. 7-6) for saturated-unsaturated water flow and advection-dispersion type equations for heat and solute transport. The water flow equation incorporates a sink term to account for water uptake by plant roots. The flow equation may also consider dual-porosity type of flow in which one fraction of water content is mobile and another fraction immobile, or dual-permeability type of flow involving two mobile regions, one representing the matrix and one the macropores. The Hydrus-1D program can be used to simulate such processes as precipitation, irrigation, infiltration, evaporation, root water uptake, soil water storage, capillary rise, deep drainage, groundwater recharge.

Specifically, the Richards' equation (Eq. 7-6) was used to simulate transient saturated/unsaturated water flow in porous medium under specific upper and lower boundary conditions:

$$\frac{\partial\theta}{\partial t} = C(h)\frac{\partial h}{\partial t} = \frac{\partial}{\partial z}\left[K(h)\left(\frac{\partial h}{\partial z}+1\right)\right] - S \tag{7-6}$$

where θ is the volumetric water content; t is the time (T); $C(h)$ is specific moisture capacity, $C(h) = \partial\theta/\partial h$, the first derivative of soil water retention curves described by van Genuchten equation (van Genuchten, 1980); h is soil water matric head (L); z is vertical coordinate (L), assumed positive upward; $K(h)$ is the unsaturated hydraulic conductivity (L T^{-1}) which is determined by van Genuchten-Mualem equation (Mualem, 1976; van Genuchten, 1980); S is a sink term which represents the volume of water removed by crop-uptake per unit time form a unit volume of soil (T^{-1}).

Moreover, the degree of water stress (DWS), which could be then utilized to calculate the relative yield according to the FAO approach (Doorenbos and Kassam, 1979) or determine the amount of water need for irrigation, can be calculated based on the simulated potential and actual crop transpiration from model's outcomes.

$$DWS = \frac{T_a}{T_x} \tag{7-7}$$

where T_a is the actual crop transpiration under specific situation as determined by the available water supply to crop; T_x is maximum crop transpiration considering that crop water requirements are fully met.

Theoretically, Hydrus-1D is not limited to any particular spatial or temporal scale, providing that the governing equations are formulated properly and can be used at that scale. However, it is not recommended to use it for large three-dimensional applications (i.e. catchment size), because the highly non-linear Richards' equation requires relatively fine spatial discretization, especially at locations where large hydraulic gradients are expected. For example, close to the soil surface, the variable meteorological conditions can cause very rapid changes in soil-water contents and corresponding pressure heads. In practice, there have been successful applications of Hydrus-1D at scales ranging from small laboratory soil columns (Yurtseven et al., 2013) to agricultural applications for soil profiles of one or several meters deep (Gärdenäs et al., 2005), up to soil profile several hundred meters deep (Scanlon et al., 2003). Particularly, for paddy soil environments, Hydrus-1D has been successfully used in simulating soil water and nutrients balances (Dash et al., 2015; Li et al., 2014; Li et al., 2015; Tan et al., 2014), and transport of water and chemical solutes (Phogat et al., 2010; Tan et al., 2015) under different management practices, i.e., direct-seeding or transplanted rice cultivation, of paddy rice cultivations.

7.2.3. Model inputs

Most soil-water models combine plant water use, initial soil water storage and water table fluctuations in varying degrees of complexity to predict soil-water storage and plant water availability. Although the major processes employed in soil water modeling are similar, the level of detail in each component varies significantly among the used models (Ranatunga et al., 2008). In general, the soil-water models require a range of meteorological, soil, crop, and land use information for representing and modeling the soil-water-plant-atmosphere interactions and processes.

Evaluation of PTFs' performance in terms of functional outcomes by using agro-hydrological model were carried out for two rice growing seasons, the so-called Dong Xuan (DX) and He Thu (HT) rice seasons. For modeling, the growing period of HT rice (simulated sowing date is 15 July, and harvesting date is 18 October) fell within the wet season (from May to November), hence rice is grown under rain-fed conditions due to the abundance of water from precipitation. The growing period of DX rice (simulated sowing date is 9 November, and harvesting date is 11 February) starts at the end of rainy season and lasts till the middle of the dry season, hence rainfall, particularly from the middle to the end of the growing period, is very limited and insufficient for rice development. DX rice is normally cultivated with supplementary irrigation. Simulations were started 7 days before sowing date to provide warm-up period for models' simulation.

7.2.3.1. Soil data

Soil data, and particularly soil hydraulic properties, are important inputs to soil-water models. As clearly stated in literature, different types of models require different levels of SWRC information for model simulation. The conceptual crop-water model AquaCrop requires SWRC information at field capacity (FC), permanent wilting point (PWP), and saturation (SAT) for the calculation of the soil water balance using a simple bucket approach, while the hydrological models Hydrus-1D needs information on the entire SWRC (e.g., van Genuchten parameters of θ_s, θ_r, α, n) for numerical solving the Richards' equation that governs water and solute movement in soil.

In this study, due to the absence of reliable measured data of hydraulic conductivity (K), we were forced to use saturated hydraulic conductivity (K_{sat}) values obtained from the PTFs of Saxton and Rawls (2006) in all simulation runs. The SWRC which was generated by three different methods (described below) was interchangeably used in the two agro-hydrological models. In other words, the functional evaluation was done by comparing simulation runs of identical K_{sat}, but different $\theta(h)$ functions.

The SWRC of the horizons in the selected profiles were generated using three different methods:

Method 1: Direct measurement of the matric head (h) - soil-water content (θ) relationship using a combination of the hanging water column and pressure plates as outlined above. A set of van Genuchten parameters (θ_s, θ_r, α, n) was then available for every soil horizon of each investigated profile (as presented in **Table 7-1**).

Method 2: Use of locally-derived PTFs to predict the SWRC, further referred to as VMD-PTFs. The best performing PTFs for soils of the Vietnamese Mekong Delta (VMD) developed with Support Vector Machines for Regression method and with the incorporation of categorical soil structural information (according to evaluated results in **Chapter 5** and **6**) were selected to that end. These PTFs are so-called point PTFs and predict in our case eight points of the θ-h relationship (at matric heads similar as in Method 1). The soil water retention parameters required in Hydrus-1D were then obtained by fitting the van Genuchten equation to PTFs' predicted θ-h data.

Method 3: Use of available PTFs which are embedded in the AquaCrop and Hydrus-1D models. In order to make the models as attractive as possible to the users, they include facilities that provide a global approximation of soil hydraulic data based on the basic soil data of the horizons (e.g. soil texture, bulk density, and organic matter content). Particularly, the users of AquaCrop model can use 'Soil Water Characteristic' software to get an estimation of SAT, FC, PWP, Ks using PTFs developed by Saxton et al. (1986) and Saxton and Rawls (2006), further referred to as SWC-PTF. Hydrus-1D users can obtain the needed sets of hydraulic parameters by using Rosetta software which was developed based on the Artificial Neural Networks PTFs of Schaap et al. (2001), further referred to as Rosetta-PTF. Our experience is that many users are using these PTFs without due consideration of their prediction accuracy and effect on the model outcomes.

7.2.3.2. Climatic data and boundary conditions of the models

In addition to soil hydraulic data, the accuracy of model simulations is also dependent on the quality of other input data. The prediction accuracy of water balance models relies on the successful characterization of climate and crop conditions prevailing throughout the simulation period (Espino et al., 1996). Daily climatic data, including maximum and minimum temperature, wind speed, humidity and daylight hours were collected from the meteorological station at Tien Giang province for the year 2011. This climatic information was utilized to calculate daily values of potential evapotranspiration (ET_o) using FAO Penman Monteith equation (Allen et al., 1998) embedded in 'ET_o Calculator' software (FAO, 2009). ET_o and the other weather data characterize the upper boundary condition, while average ground water level normally observed in the field during wet and dry seasons, i.e., 0.8 m and 1.2 m, respectively (Le, 2003), was used as fixed lower boundary condition for model simulations. The same climatic information and boundary conditions were constantly applied in all simulated scenarios (i.e., for the four investigated soil profiles with three simulation runs corresponding to three soil hydraulic data sets obtained from different methods) of two agro-hydrological models with the assumption that not much variation in climatic condition was present in the entire study area. The daily information of rainfall and ET_o of the two investigated crop seasons is graphically presented in **Figure 7-1**.

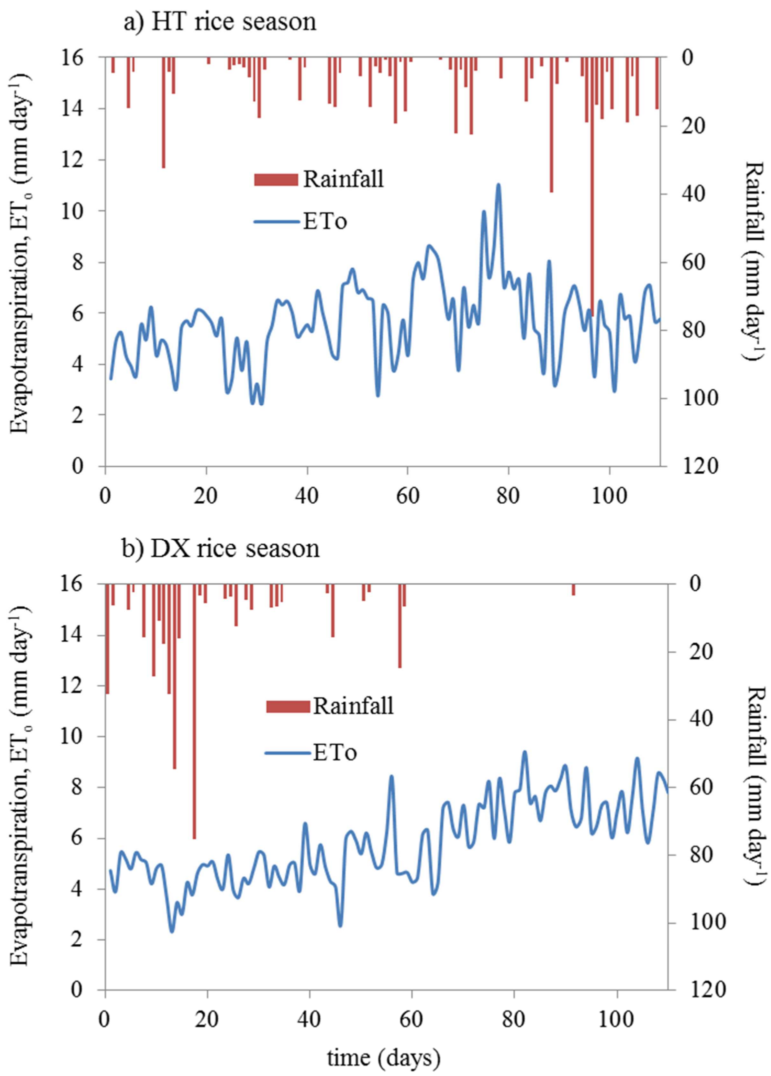

Figure 7-1. Daily precipitation and reference evapotranspiration ET_o calculated from FAO Penman Monteith equation for the two investigated rice growing seasons in 2011, i.e. (a) He Thu (HT) and (b) Dong Xuan (DX), with meteorological data from the meteorological station at Tien Giang province, Vietnam.

7.2.3.3. Crop data

Short-length duration rice varieties (95-110 days) which are normally cultivated in the region were selected for model simulation. For the AquaCrop model, several conservative crop parameters of paddy rice which were already validated and available in the AquaCrop's crop module were utilized with some modification based on rice crop information from local observations (Nguyen, 2011).

In the Hydrus-1D model, the actual rice root uptake, the sink term S, in Richards' equation was estimated using the Feddes model (Feddes et al., 1978):

$$S(h) = \alpha(h) * Sp \qquad (7\text{-}8)$$

where *Sp* is the potential water uptake rate (cm day^{-1}) and *α(h)* is a dimensionless function of the soil water pressure head in response to water stress, taking values between 0 to 1. Feddes et al. (1978) proposed a piecewise linear reduction function parameterized by four critical values of the water pressure head, *h4* < *h3* < *h2* < *h1* to describe water stress. The values of Feddes' parameters for paddy rice in this study were based on Phogat et al. (2010) with *h1* = 100 cm, *h2*= 55 cm, *h3L*= -160 cm, *h3H* = -250 cm, *h4*= -15000 cm. L and H denote low and high evaporative demand, respectively.

7.2.4. Statistical evaluation of simulated time series

It is worth noting that a 'true' evaluation of the model outcomes and PTF performance was not performed in this study. The functional evaluation conducted here was based on the assumption that simulations using lab-measured hydraulic data serve as benchmark for comparison since it is the most feasible and accurate soil data that is at stake for handling the large scale model simulation. In a study with highly varying soil-water contents, Rezaei et al. (2016) showed that Hydrus-1D was well capable of simulating trends in soil-water content change using a forward procedure based on lab determined soil hydraulic properties.

For AquaCrop simulation, when a unique value of model outcomes (e.g., rice yield, net irrigation requirement, soil water balance components) was reported, the evaluation was performed by comparing the deviation of the outcomes between two indirect methods (i.e. VMD-PTFs and SWC-PTFs) to those of the lab-measurement scenario. When time series of functional criteria (e.g., soil-water content in the root zone or at different depth, and degree of water stress) were presented (in both AquaCrop and Hydrus-1D simulations), the performance of the different data-derived approaches was quantitatively compared by using several statistical criteria such as mean error (ME), standard deviation of error (SDE), and root mean square of error (RMSE)

$$ME = \frac{1}{N}\sum_{i=1}^{N}\left(E_i - M_i\right) \qquad (7\text{-}9)$$

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(E_i - M_i\right)^2} \qquad (7\text{-}10)$$

$$SDE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(e_i - \bar{e}\right)^2} \qquad (7\text{-}11)$$

where E_i is the i^{th} simulation using PTFs' estimated SWRC data, M_i is the i^{th} simulation using measured SWRC data, e_i is the i^{th} error of time series variable which was determined by subtracting simulations using measured data to those using PTFs' estimated SWRC data, $\bar{e}$ is the mean of the time series errors, N is the number of units of time-series.

7.3. Results and Discussions

7.3.1. Comparison of SWRC

The SWRCs of the four studied soil profiles obtained by the three methods, i.e., M1, lab derived SWRC; M2, using VMD-PTFs; and M3, using PTFs embedded in the AquaCrop (SWC-PTFs) and Hydrus-1D (Rosetta PTFs) models, are graphically displayed in **Figure 7-2**, and the parameters as needed in the respective models are presented in **Table 7-2** and **Table 7-3**. The continuous curves of M1 and M2 were obtained by fitting the van Genuchten equation to lab-measured and VMD-PTF predicted points of SWRC.

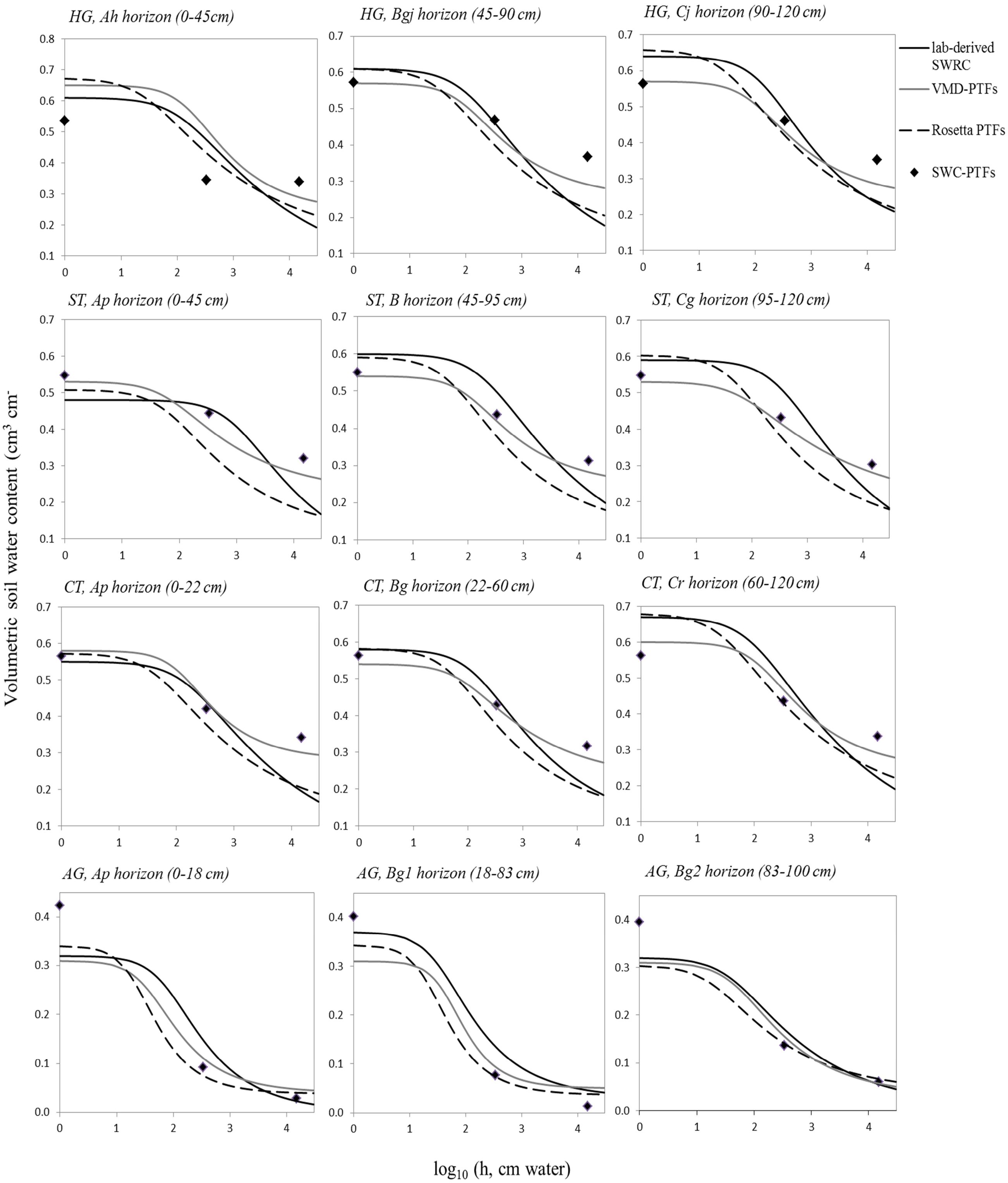

Figure 7-2. Soil water retention characteristics (SWRC) of three horizons of four studied profiles obtained from direct laboratory measurement (lab-derived SWRC) and indirectly predicted using locally-derived PTFs (VMD-PTFs), or globally provided PTFs (SWC-PTFs and Rosetta-PTFs).

Table 7-2. Soil hydraulic inputs obtained from VMD-PTFs and SWC-PTFs of Saxton and Rawls (2006) for AquaCrop model simulation.

Location	horizon	VMD-PTFs					SWC-PTFs (Saxton and Rawls, 2006)					
		FC (m^3 m^{-3})	PWP (m^3 m^{-3})	SAT (m^3 m^{-3})	AWC (m^3 m^{-3})	DW (m^3 m^{-3})	FC (m^3 m^{-3})	PWP (m^3 m^{-3})	SAT (m^3 m^{-3})	AWC (m^3 m^{-3})	DW (m^3 m^{-3})	K_{sat} (cm day^{-1})
HG	Ah	0.50	0.29	0.65	0.21	0.15	0.35	0.34	0.54	0.01	0.19	32.2
	Bgj	0.44	0.29	0.57	0.15	0.13	0.47	0.37	0.57	0.10	0.10	6.00
	Cj	0.44	0.29	0.57	0.15	0.13	0.46	0.35	0.56	0.11	0.10	6.06
ST	Ap	0.41	0.28	0.53	0.13	0.12	0.44	0.32	0.55	0.12	0.10	6.33
	Ab	0.42	0.28	0.54	0.14	0.12	0.44	0.31	0.55	0.12	0.11	8.06
	Cg	0.42	0.28	0.53	0.14	0.11	0.43	0.30	0.55	0.13	0.11	8.44
CT	Ap	0.44	0.30	0.58	0.14	0.14	0.42	0.34	0.57	0.08	0.15	15.9
	Bg	0.43	0.29	0.54	0.14	0.11	0.43	0.32	0.56	0.11	0.13	13.2
	Cr	0.46	0.29	0.60	0.17	0.14	0.44	0.34	0.56	0.10	0.13	10.8
AG	Ap	0.11	0.05	0.31	0.06	0.20	0.09	0.03	0.42	0.06	0.33	237.6
	Bg1	0.10	0.05	0.31	0.05	0.21	0.08	0.01	0.40	0.06	0.33	271.5
	Bg2	0.16	0.06	0.31	0.10	0.15	0.14	0.06	0.40	0.08	0.26	107.7

FC is the soil water content (m^3 m^{-3}) at Field Capacity (represent to matric potential of -33 kPa for fine-textured soils, and -10 kPa for coarse-textured soils;

PWP is the soil water content (m^3 m^{-3}) at Permanent Wilting Point (represent to matric potential of -1500 kPa);

SAT is the soil water content (m^3 m^{-3}) at saturation (we take the values of θ_s from fitted vG parameters);

AWC is the total available water content (m^3 m^{-3}), *AWC = FC – PWP*;

DW is the drainable water content (m^3 m^{-3}), *DW = SAT – FC*;

K_{sat} is saturated hydraulic conductivity (cm day^{-1}) obtained from Saxton and Rawls (2006) PTFs in the ‘Soil Water Characteristics’ software.

Table 7-3. Soil hydraulic parameters obtained via fitting the equation to point predicted SWRC using VMD-PTFs and via parameter-based PTFs of Schaap et al. (2001) in Rosetta software for Hydrus-1D model simulation.

Location	Horizon	Fitted parameters of point predicted SWRC of VMD-PTFs				Predicted parameters from Rosetta PTFs			
		θ_r	θ_s	α	n	θ_r	θ_s	α	n
HG	Ah	0.24	0.65	0.006	1.47	0.12	0.67	0.034	1.23
	Bgj	0.25	0.57	0.011	1.40	0.12	0.61	0.025	1.26
	Cj	0.24	0.57	0.011	1.39	0.12	0.66	0.030	1.25
ST	Ap	0.22	0.53	0.015	1.32	0.10	0.51	0.016	1.31
	Ab	0.24	0.54	0.010	1.39	0.11	0.59	0.020	1.30
	Cg	0.18	0.53	0.015	1.23	0.11	0.60	0.020	1.31
CT	Ap	0.28	0.58	0.008	1.56	0.11	0.57	0.021	1.28
	Bg	0.22	0.54	0.012	1.31	0.11	0.58	0.020	1.30
	Cr	0.24	0.60	0.008	1.41	0.12	0.68	0.032	1.25
AG	Ap	0.04	0.31	0.027	1.60	0.04	0.34	0.046	1.76
	Bg1	0.05	0.31	0.022	1.88	0.04	0.34	0.048	1.75
	Bg2	0.03	0.31	0.020	1.41	0.04	0.30	0.050	1.33

When comparing the lab-derived SWRC (**Table 7-1**) with those estimated by using PTFs, discrepancies can be observed among the studied profiles and the used PTFs. For the acid sulfate soil profile with distinctively high organic matter content in the surface horizon (HG profile), taking lab-measured data as a reference, VMD-PTFs offered a more accurate estimation of SWRC at SAT, FC and PWP in the surface horizon (with deviations of 0.01 to 0.04 m^3 m^{-3}) than the SWC-PTFs (deviations are about -0.14 to 0.09 m^3 m^{-3}). As a result of the systematic over-estimations of VMD-PTFs at the three hydraulic points, the plant extractable water content (so-called available water content, AWC) and drainable water content (DW), two other important soil hydraulic properties in AquaCrop, calculated from VMD-PTFs did well agree with those from lab-derived data with deviations of about 0.03 m^3 m^{-3}. In lower horizons, both PTFs (VMD-PTFs and SWC-PTFs) under-estimated SWRC at SAT and FC but over-estimated PWP, in which much larger deviation was observed with SWC-PTFs. Because of that, the calculated AWC and DW based on SWC-PTFs were much more different from those calculated by lab-derived data as compared to those from VMD-PTFs.

As regards the prediction of van Genuchten parameters, the VMD-PTFs outperformed the Rosetta PTFs in depicting the SWRCs of the surface horizon. Although there was an over-estimation of θ_r, θ_s, and n parameters, the values of the shape parameter α were comparable between the two approaches (i.e. lab-derived data and VMD-PTFs), hence leading to analogous shapes of SWRC as observed in **Figure 7-2**. In the subsurface horizons, although there was still a good match of SWRC up to matric head of -1000 cm ($\log_{10}$(h) of 3), the dry range of the curve was not well compatible as θ_r predicted by VMD-PTFs was higher than that of lab-derived data. Much variation between the SWRCs derived by curve-fitting of lab-derived data and VMD-PTFs was witnessed in the third horizon of HG profile. On the other hand, using the parameter-based Rosetta-PTFs, the SWRCs of all three horizons had comparable matches to the ones of lab-derived data, particularly at two ends of the curve as they all have close values of θ_r, θ_s, and n. However, the shape parameter α of Rosetta-PTFs was far higher than that of based on the lab-measurements which indicates a more sudden change in soil water content under small change of negative matric heads, so large deviation to measured data was observed in the intermediate range of the curve.

In case of the salt-affected soil (ST profile), rather large differences between PTF estimations (i.e., VMD-PTFs and SWC-PTFs) and the lab-derived SWRC were observed, particularly in the surface horizon. Both indirect PTFs methods provided similar prediction of SWRC at SAT, FC and PWP which were far away from the lab-measured data. Specifically, VMD-PTFs and SWC-PTFs overestimated SAT of topsoil and PWP of all horizons, whereas they underestimated FC. As a result of that, calculated AWC and DW from VMD-PTFs and SWC-PTFs were much different to those from lab-derived data. Similar to point estimation, comparable outcomes of SWRC parameters (e.g., θ_r, θ_s, and n) were obtained with PTF methods. The α values from lab-derived data were much smaller (from 0.001 to 0.005) as manifested by the long stable plateau up to $\log_{10}$(h) of 2.3 in the surface horizon and up to $\log_{10}$(h) of 2 for the two subsoil horizons (**Figure 7-2**), while those of VMD-PTFs and Rosetta-PTFs displayed somehow greater values (e.g., from 0.010 to 0.015, and from 0.016 to 0.020 for VMD-PTFs and Rosetta-PTFs, respectively). It is well known that swelling and dispersion of soil colloids is affected by the composition of the clays' exchangeable cations. Adsorbed sodium ions form a wide diffuse double layer, creating high swelling pressures, and form single clay platelets which tend to persist in dilute solution. Swelling and/or dispersion of soil colloids alters the geometry of soil pores and thus affects the soil's hydraulic properties. Thus, salt-affected soils cannot be considered having an inert porous medium whose hydraulic properties can be predicted from its texture or pore size distribution (Shainberg and Letey, 1984). Moreover, as it was already mentioned in section 7.2.1, the surface horizon of the ST profile was fully disturbed due to the effect of soil preparation for upland crop cultivation. Soil management, together with salinity effect, significantly altered the architecture of pore system on soil surface of the ST profile. As a consequence, the hydraulic behavior of surface horizon which is directly subjected to wetting and drying processes was hardly described by the indirect PTF approaches.

For the recently-developed alluvial soil with a certain stratification in soil texture (CT profile), VMD-PTFs performed well in predicting of FC (deviations of 0.01-0.03 m^3 m^{-3}) and SAT (error equals 0.03-0.07 m^3 m^{-3}), but less accurate for PWP (deviations are around 0.07-0.11 m^3 m^{-3}). Similarly, SWC-PTFs also provided rather good estimations of FC and SAT (except that of the third horizon), but did not predict well PWP. Due to the larger deviations in the prediction of PWP (in the range of 0.11-0.15 m^3 m^{-3}), the AWCs calculated with SWC-PTFs' scenario were further away from measured ones as compared with those of VMD-PTFs. For the whole SWRC estimation, graphically, the curves of VMD-PTFs matched well to those of lab-measured data in the intermediate range, but deviated more at the two ends, particularly at the dry end of the curve, whereas the curve of Rosetta-PTFs was closer to lab-measured SWRC at the two ends, but deviated more in the middle of the curve (α value of Rosetta-PTFs are three times higher than those obtained from data of lab-measurement).

Finally, for the coarse-textured soils in the mountainous area (AG profile), the outperformance of the VMD-PTFs is rather evident. Specifically, they predicted well SWRC at FC, PWP and SAT with the largest error of 0.06 m^3 m^{-3} being observed for SAT in the second horizon. Although SWC-PTFs predicted FC and PWP adequately, those PTFs did not perform well in estimation of SAT. As a result of that, a large deviation of DW calculated from SWC-PTFs is noticed as compared to those from lab-measured data. For continuous SWRC estimations, **Figure 7-2** displays rather analogous curves obtained from lab-derived SWRC and VMD-PTFs, particularly for the first and third horizons. The Rosetta PTFs also performed well, but showed a little more deviation compared to VMD-PTFs due to higher predicted values of shape parameters *α* and *n*.

Generally, for the estimation of SWRC, the locally derived VMD-PTFs outperformed the PTFs embedded in the models' facilities in several cases. These results are in well accordance to the findings presented by Tranter et al. (2009). These authors concluded that models which are trained on large and diverse datasets offer larger domain and greater coverage, but are far less precise than those trained on less diverse data. Additionally, it is possibly true that the hydraulic behavior of some major soil groups found in VMD (e.g., alluvial soils and acid sulfate soils of heavy texture and high organic matter content) are not well represented by globally-offered PTFs which are predominantly trained on soil databases from temperate climates as we already discussed in **Chapter 3**. Moreover, since VMD-PTFs were developed based on SVR in which categorical soil structure information was incorporated, their proved outperformance for coarse textured soils (results of **Chapter 6**) is once again demonstrated on the AG profile in this evaluation study. The observed worse performance of the PTFs on the salt-affected soil (ST profile) is explainable because of the dynamic soil-water retention behavior of saline soils (Shainberg and Letey, 1984), which could apparently not be captured correctly by the empirical relationships developed from international databases and a local VMD data set. Moreover, as the local training data was collected from a cross-sectional study design, locally-derived PTFs have a limitation in explaining temporal variability of soil properties in the studied region (Kutner et al., 2005), which in turn could probably be an additional erroneous source for the prediction using VMD PTFs.

Since there are obvious deviations in SWRC directly determined by lab-measurement, and indirectly predicted via locally derived VMD-PTFs or via globally used PTFs embedded in models, it is important to assess to what extent the observable differences in soil hydraulic data propagate to the outputs of agro-hydrological models. The functional evaluation displayed below will provide important answers for practical applications of PTFs.

7.3.2. AquaCrop model

Using AquaCrop, the performance of different methods used to determine SWRC was evaluated in terms of attainable rice yield and daily variation of root-zone water storage in the wet season, net irrigation requirement for rice cultivated in dry season, and soil water balance components in both seasons.

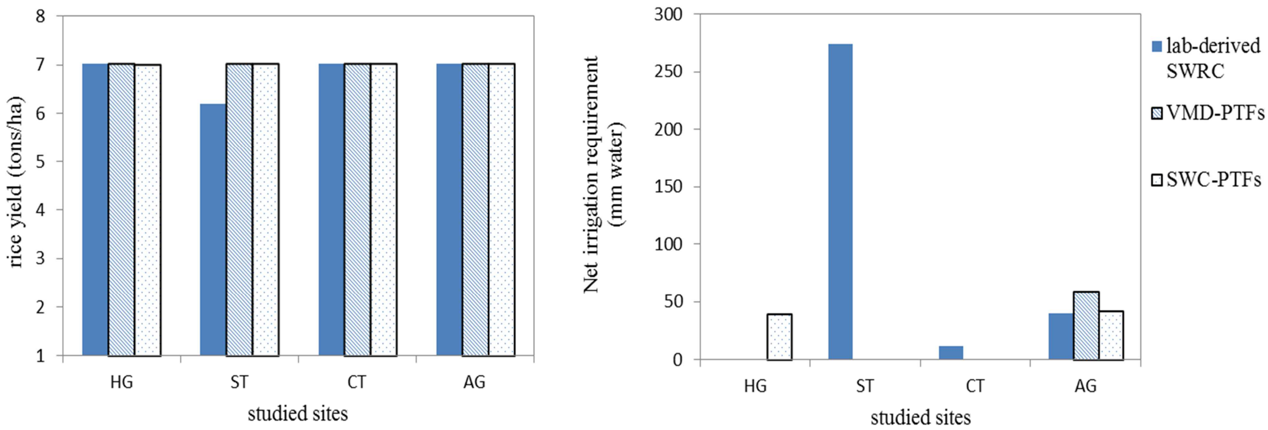

Figure 7-3. Simulated rice yield in wet season and net irrigation requirement in dry season determined by AquaCrop model using SWRC data derived from different approaches (i.e., lab-derived SWRC, prediction via using VMD-PTFs and SWC-PTFs).

As it is clearly observed in **Figure 7-3**, in three of the four investigated profiles, the simulated rice yield under rain-fed condition of wet season was almost the same in the three simulation scenarios that used different soil hydraulic data sets (rice yield was around 7 tons ha^{-1} in all scenarios). The exception was ST profile where rice yield obtained with lab-measured data (6.2 tons ha^{-1}) was remarkably lower than those of VMD-PTFs and SWC-PTFs (around 7 tons ha^{-1}).

Similarly, in the dry season when sole precipitation, particularly in the half-end period of the growing season, is not sufficient to meet the crop demand (**Figure 7-3**), supplemental irrigation is needed to ensure optimal moisture conditions for crop development and yield formation. The net irrigation requirements calculated by using three different sets of hydraulic data displayed slight variation for HG, CT and AG profiles, while a noticeable difference was recorded for the ST profile (**Figure 7-3**). Specifically, in case of HG profile, no irrigation was needed for rice in the dry season under the scenarios of using lab-derived data and VMD-PTFs data, while there was around 40 mm of irrigation water needed for optimal rice growth when data of SWC-PTFs was used. For the alluvial soil profile (CT), using lab-derived data as input resulted in 12 mm of irrigation requirement, while PTFs' predicted inputs (i.e., VMD-PTFs and SWC-PTFs) came up with no irrigation requirement. Similarly, 40, 59, and 42 mm irrigation was proposed for rice cultivated in loamy sand soil of the AG profile, when soil input data was obtained from lab-measurement, VMD-PTFs, and SWC-PTFs, respectively. In case of the ST profile, using lab-derived data ended up in 274 mm water for net irrigation, while using VMD-PTFs and SWC-PTFs data revealed that no irrigation was needed for rice in dry season.

Since AquaCrop is a water-driven model in which attainable yield was calculated on the basis of soil water productivity (which is adjusted to daily water content), the difference in simulated yield as well as the amount of water required to obtain full crop development should be investigated in accordance with simulated soil water balance components (expressed in **Table 7-4**).

Indeed, the root zone can schematically be considered as a reservoir whose capacity is determined by total available water content (AWC = FC - PWP). By keeping track of the incoming and outgoing water fluxes at the boundaries of the root zone, the amount of water retained in the root zone can be calculated by means of a soil water balance. Water is added to the soil reservoir by rainfall, irrigation, and capillary rise from shallow ground water table, and removed from the soil pool by the processes of soil evaporation, crop transpiration, and deep percolation.

As it is visually displayed in

Figure 7-4, the fluctuation of soil water storage (in terms of trend and absolute quantity) was well agreed between simulations using lab-derived SWRC and VMD-PTFs predicted SWRC in CT and HG profiles. These simulations were rather deviated from the scenario using SWC-PTFs predicted data. Inversely, in case of salt-affected soil in ST profile, similar daily changes of soil water content were observed between PTFs predicted scenarios (i.e. VMD-PTFs and SWC-PTFs) which in turn substantially different from that of lab-derived SWRC simulation. These results are highly correspond to simulation of attainable rice yield in rainy season as manifested in **Figure 7-3**. On the other hand, no clear correlation trends of soil water storage fluctuation were observed among simulation scenarios in AG profile. That might be explained by the discrepancy of SWRC at FC and PWP predicted by VMD-

PTFs and SWC-PTFs from those of laboratory measurement. Such difference leads to the variation in calculating soil water storage capacity of sandy soil in AG profile.

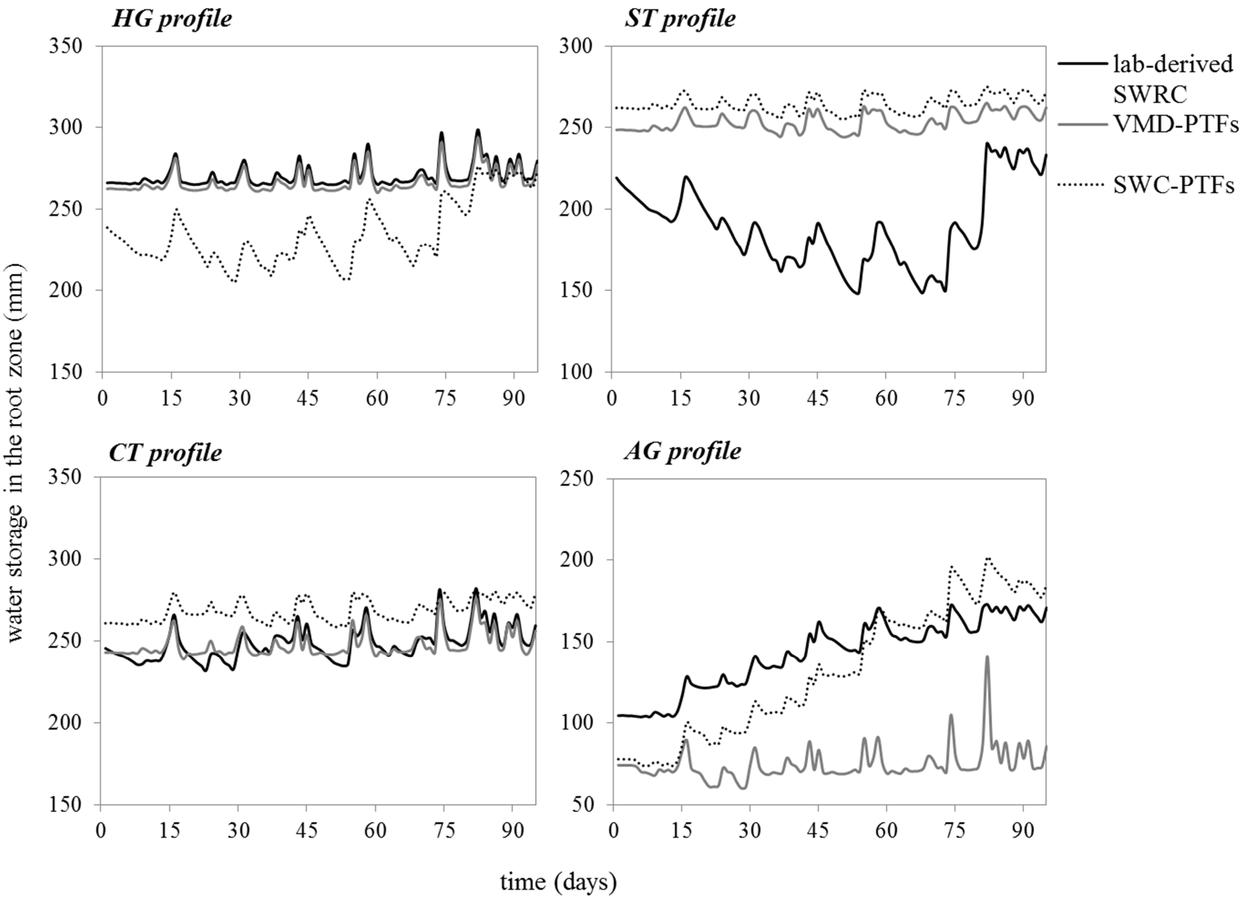

Figure 7-4. Daily variation of soil water storage (mm) in the root zone (Z_r = 0.5m) in the wet season of four studied profiles under different scenarios of SWRC inputs.

Table 7-4. Outputs of soil water balance components obtained by simulations using inputs of (M1) lab-measurement, (M2) VMD-PTFs' prediction, and (M3) SWC-PTFs' prediction.

Site	Evaporation (mm)			Crop transpiration (mm)			Drainage (mm)			Capillary Rises (mm)			Change in soil water storage (mm)		
	M1	M2	M3	M1	M2	M3	M1	M2	M3	M1	M2	M3	M1	M2	M3
	HT rice season (rainy season)														
HG	150	150	106	374	374	373	330	327	144	292	290	92	13	14	40
ST	142	148	148	322	374	374	112	250	256	6	232	236	6	30	28
CT	133	149	147	374	374	374	268	329	246	210	288	232	10	12	35
AG	139	103	110	374	374	374	211	321	173	237	237	220	71	13	111
	DX rice season (dry season)														
HG	128	130	117	345	345	345	220	220	194	337	339	240	1.6	1.6	-20
ST	128	125	125	343	345	345	192	209	209	13	323	323	-18	1.6	1.5
CT	121	129	128	345	345	345	191	214	216	264	333	333	-25	1.5	1.4
AG	114	101	97	345	345	345	218	238	230	277	266	274	-3.5	-2.8	-0.7

Additionally, as it is clearly shown in **Table 7-4**, the simulated soil water components in the HG profile (i.e., soil evaporation, crop transpiration, deep drainage, capillary rise and change in water storage in the root zone) obtained with the lab-measured data (M1) and VMD-PTFs' predicted data (M2) scenarios were almost similar, whereas much different values (e.g., smaller or larger) were obtained under the SWC-PTFs scenario (M3). A similar trend was observed in the dry season though less pronounced, due to the difference in climatic conditions between the two seasons, as well as to additional irrigation water being added to meet the crop demand in the dry season. Inversely, for the ST profile, comparable outcomes were observed with M2 and M3 for all above-mentioned soil water balance components, while those of M1 were much smaller in both rice growing seasons. In case of the CT and AG profiles, no clear trend in the fluctuation among soil water balance components was manifested under the different simulation scenarios. Using various soil hydraulic data sets lead to slight variation in simulated outcomes of soil water balance components. Such slight variation together with the compensation between incoming and outgoing fluxes to the root zone (e.g., little water loss by drainage and soil evaporation was paid off by less water moving upward from the groundwater table), resulted in more or less similar simulated rice yield and net irrigation requirement for the CT and AG profiles (**Figure 7-3**).

It is important to note that accurate estimation of AWC in AquaCrop is probably as important as precise estimation of SWRC at FC, PWP and SAT, since AWC determines the capacity of the soil reservoir in retaining and transporting the water as well as the limited threshold of allowable soil water depletion. **Table 7-1** and **Table 7-2** present the wide range of AWC calculated from direct laboratory measurement and indirect PTFs estimations of SWRC among the studied soil profiles. As it is demonstrated in the outcomes of the AquaCrop model for the HG profile, deviation of soil water storage in the root-zone (

Figure 7-4) and soil water balance components (**Table 7-4**) between simulated scenarios are closely determined by the error of calculated AWC, besides that of FC, PWP and SAT estimations. Indeed, by investigating the reliability of different indirect methods to estimate SWRC as well as the error propagation of SWRC estimation through crop models (i.e., ACRU and CROPGRO Soybean model), Leenhardt (1995) and Gijsman et al. (2003) concluded that a crop-water model is not a simple linear model based on FC and PWP, but a more complex one where the retention and movement of water in the soil profiles are determined by these properties and the difference between them (i.e. AWC). They also warned users of crop-water models in that they must be concerned not only with the reliability of the model itself and its sensitivity to changes in soil parameters, but also with the reliability of indirect methods available for soil data estimation.

7.3.3. Hydrus 1D model

Using the physically-based Hydrus-1D model, the performance of PTFs was evaluated in terms of soil-water content distribution in the root zone in the wet season, and the degree of water stress (defined as actual crop transpiration over the potential one) in the dry season (presented in **Figure 7-5** and **Figure 7-6**). Model simulations were carried out using soil hydraulic parameters derived by direct laboratory measurement and PTF estimations (i.e., VMD-PTFs and Rosetta PTFs).

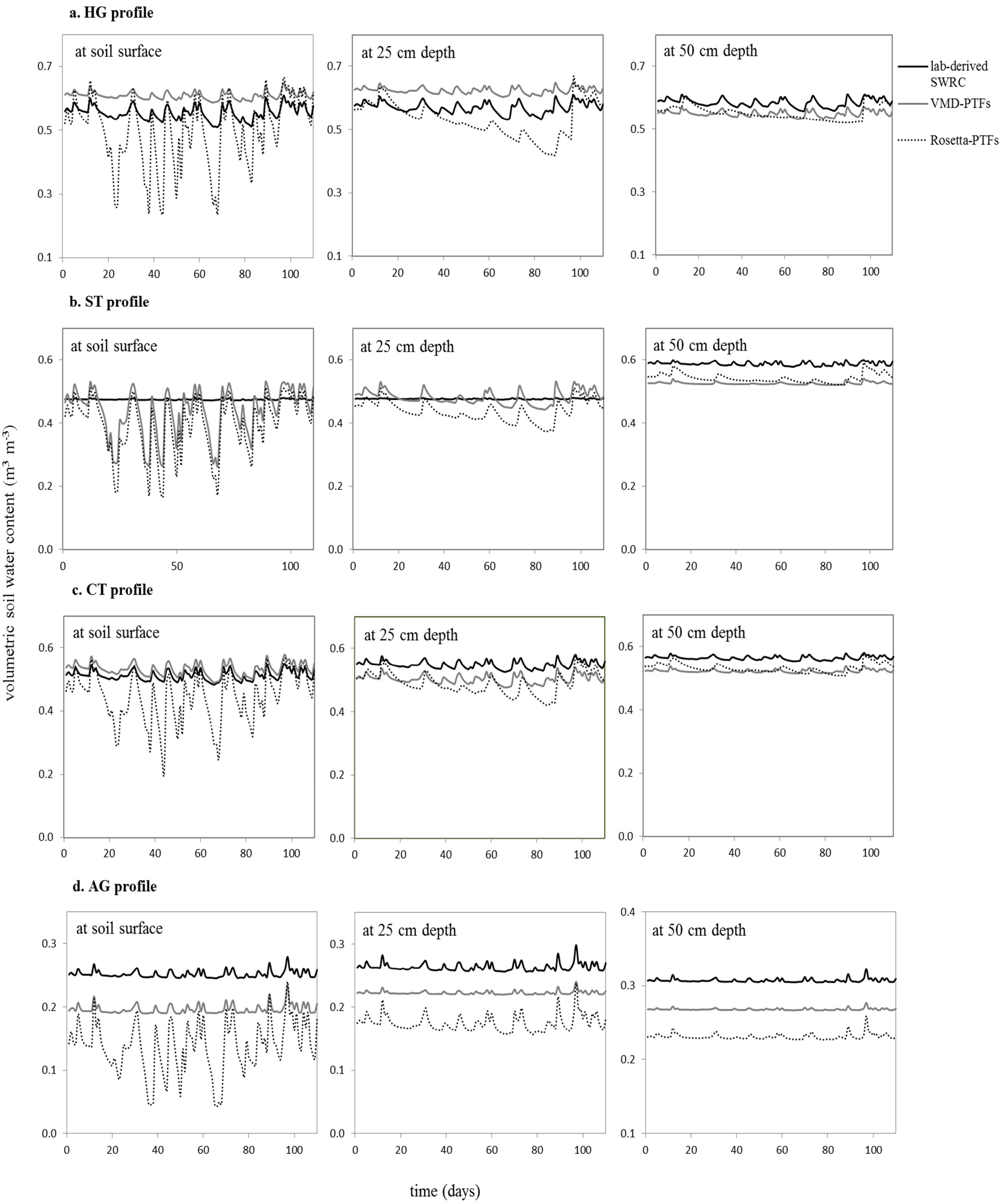

Figure 7-5. Simulated daily moisture content distribution at three depths within the root-zone of four studied profiles under different simulation scenarios of using SWRC data determined by lab measurement (lab-derived SWRC) and via prediction with locally derived PTFs (VMD-PTFs) and globally offered PTFs (Rosetta-PTFs).

Table 7-5. The mean error (ME) of time series soil water content (m^3 m^{-3}) simulated by using predicted data of VMD-PTFs and Rosetta-PTFs where simulation using lab-measured SWRC data was considered as benchmark for comparison. Numbers between the brackets represent the standard deviation of the error (SDE).

Studied profile	at the surface		at 25 cm depth		at 50 cm depth	
	VMD-PTFs	Rosetta-PTFs	VMD-PTFs	Rosetta-PTFs	VMD-PTFs	Rosetta-PTFs
HG	0.054 (*0.011*)	-0.06 (*0.089*)	0.054 (*0.009*)	-0.037 (*0.051*)	-0.035 (*0.006*)	-0.032 (*0.02*)
ST	-0.039 (*0.075*)	-0.081 (*0.088*)	0.005 (*0.021*)	-0.038 (*0.032*)	-0.062 (*0.005*)	-0.045 (*0.014*)
CT	0.02 (*0.005*)	-0.07 (*0.066*)	-0.045 (*0.002*)	-0.059 (*0.03*)	-0.042 (*0.001*)	-0.032 (*0.014*)
AG	-0.055 (*0.002*)	-0.114 (*0.037*)	-0.041 (*0.004*)	-0.089 (*0.008)*	-0.039 (*0.001*)	-0.076 (*0.002*)

Upon visual inspection of **Figure 7-5**, it is obvious that in three of the four studied profiles (i.e., HG, CT, AG), the simulations of soil-water content during the wet season using VMD-PTFs (M2) were generally in good agreement with those of scenario using lab-derived data (M1) in terms of the absolute values of deviation and the fluctuation pattern of soil-water content with time during the growing season. Specifically, for the HG profile, the scenario of using VMD-PTFs over-predicted soil-water content with an average value of 0.05 cm^3 cm^{-3} at the soil surface and at 25 cm depth, while a less pronounced under-estimation of 0.04 cm^3 cm^{-3} was noticed at 50 cm depth. Besides that, the fluctuation pattern of soil water content in HG profile under M2 scenario was similar to that of M1 as expressed by the small SDE ranging from 0.006 to 0.01 for three investigated depths. Similarly, the soil water content simulations of the CT profile clearly displayed similar fluctuation patterns between runs using parameter from M1 and M2 scenarios (SDE are in the range of 0.001 - 0.005) in which using M2 slightly over-estimated soil water content at the surface (ME equals 0.02 cm^3 cm^{-3}), while under-estimated soil water content at 25cm and 50 cm depth (ME are of -0.04 cm^3 cm^{-3}). For the coarse textured soil of the AG profile, likely trends in the distribution of soil water content were also noticed between M1 and M2 scenarios (SDE are around 0.001-0.004) with absolute differences in soil water content (ME) in the range of 0.04-0.05 cm^3 cm^{-3} for the three investigated depths within the root zone. Inversely, the performance of the Rosetta PTFs (M3), was always worst in terms of absolute magnitude of variation as well as in the similarity in the patterns of time series distribution of soil water content (i.e., higher

ME values together with higher SDE of runs using Rosetta-PTFs for HG, CT and AG profiles).

There are somehow opposite observations for the ST profile where the simulations using measured SWRC data were obviously apart from those of using PTFs predicted information (i.e. M2 and M3). Indeed, the significant discrepancies in the shape of SWRCs determined by different methods (**Figure 7-2** and **Table 7-3**) resulted in an extreme difference in the response of soil hydraulic behavior to the upper meteorological conditions. Understandably, the small value of the α parameters of SWRC obtained by lab-derived SWRC reveals that a large change in pressure head associated with the upper boundary (weather) conditions do not cause a significant change in soil water content. This characteristic was manifested in the minor changes in soil water content at the different depths of the lab-measured data scenario (**Figure 7-5**). On the other hand, higher values of α obtained by PTFs methods (VMD-PTFs and Rosetta-PTFs) showed a prompt response of soil hydraulic behavior in terms of changing soil-water content under the imposition of upper climatic conditions. As a result of that, large ME and large SDE values were observed for both PTFs methods as compared to the simulations using measured data.

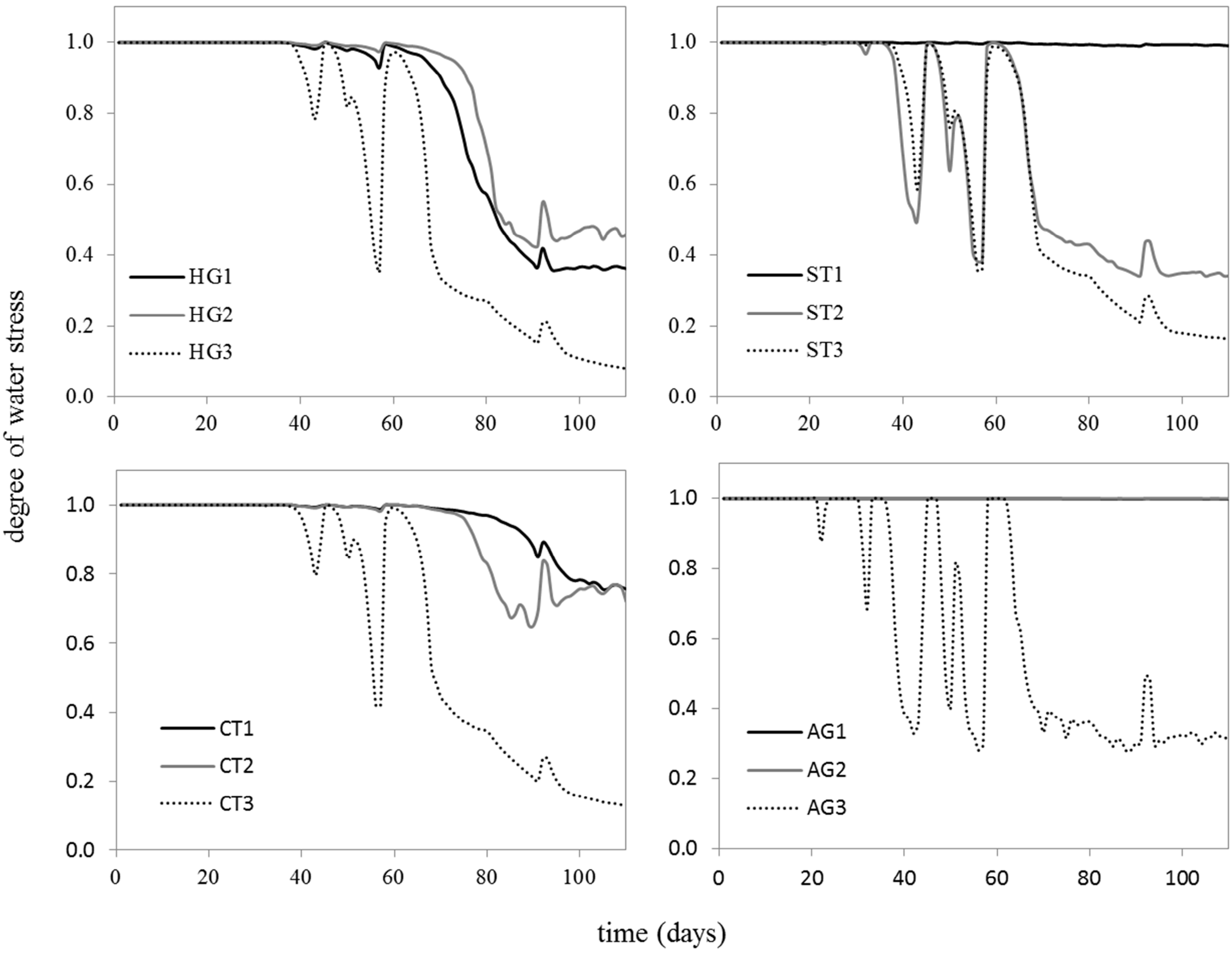

Figure 7-6. Simulated degree of water stress (DWS) in dry season of four studied profiles under different scenarios of SWRC inputs. For example: ST1, ST2, ST3 are the simulated DWS of ST profile when using SWRC parameters obtained from lab-measurement, VMD-PTFs and Rosetta-PTFs predictions, respectively.

In the dry season, the degree of water stress (DWS), i.e. actual root water uptake over the potential crop transpiration, was simulated to investigate the propagation of errors in determining soil hydraulic parameters with the three methods to the outcomes of the hydrological model (**Figure 7-6**). When DWS falls below unity, crops start suffering from a certain degree of water stress. DWS is thus an indicator to guide irrigation. As we already reported, in nearly haft end growing period in the dry season, precipitation is rare and not sufficient for crop water demand as clearly demonstrated in **Figure 7-6** where the deviation between the three simulated scenarios is observed from day 36 onwards.

In three of the four studied profiles (i.e. HG, CT, and AG profiles), the simulations of the DWS in the dry season with VMD-PTFs' data (i.e., HG2, CT2, AG2) perform generally well as their representing DWS curves were closer to the ones using lab-measured SWRC data (HG1, CT1, AG1) than those with Rosetta PTFs (HG3, CT3, and AG3). The average deviation of DWS as expressed by root mean square error (RMSE) were 0.068 for HG2, 0.23 for HG3; 0.077 for CT2, 0.4 for CT3; 0.0006 for AG2, and 0.48 for AG3. The possible explanation for such discrepancies could probably be the variation in the prediction of soil water content in the root zone among different data derived approaches (as manifested in the results of the wet season). As we have clarified, the shape parameter α of SWRC predicted by Rosetta-PTFs were higher than the corresponding ones obtained from fitting the van Genuchten equation to lab-measured SWRC data and VMD-PTFs predicted data (**Figure 7-2** and **Table 7-3**). Such large α values display a sudden change in soil water content under a small change of soil matric potential driven by atmospheric evaporative demand. These features probably explain the severe water stress observed with the Rosetta-PTFs scenario. Specifically, in the AG profile, the simulated DWS of AG1 and AG2 were almost the same and no water stress occurred in the dry season. Inspecting other soil water balance outputs (results not shown) could probably clarify for DWS results. Indeed, the simulated upward fluxes (capillary rise) from the ground water table of AG1 and AG2 scenarios were sufficiently high and can partly supply to water to address the crop transpiration demand, while it was small and not sufficient in case of AG3. Numerically, the total upward fluxes at the lower boundary in the whole simulated period were approximately 15 cm for AG1 and 16 cm for AG2 (as simulated with Hydrus-1D), while a total downward flux of 14 cm was simulated for the AG3 scenario. The resemblance in the shape of SWRCs, particularly in the first and third horizons, might possibly explain for similar soil behaviors of AG1 and AG2 in simulating DWS.

Following the same token with previous model simulations, the DWS of the ST profile simulated by using PTF predicted SWRC data (i.e., ST2 and ST3 corresponding to the simulations using predicted SWRC from VMD-PTFs and Rosetta-PTFs) was extremely different from the one of ST1 (i.e., scenario using parameters from lab-measured SWRC data). The RMSE was equal to 0.41 and 0.48 for ST2 and ST3, respectively. These results are expected as we also observed a large discrepancy in predicting soil water content in the root zone in the wet season between ST1 on the one hand and ST2 an ST3 on the other hand.

Generally, the evaluation of simulation results of AquaCrop and Hydrus-1D models once again demonstrate the need of locally-derived VMD-PTFs. Indeed, as the locally-derived PTFs resulted in a more accurate estimation of SWRC, they seem to be an attractive alternative to lab-derived labour intensive and expensive data in modeling soil water balance and crop response to water. However, for salt-affected soils, VMD-PTFs reveal a limitation in describing the dynamic SWRC of such specific soils, which in turn requires further research to improve their predictive performance. Several directions for PTF improvement can be taken into consideration, for example, including other predictor variables that determine salinity (e.g., EC) or reflect the dynamic soil water characteristics due to swelling and dispersion processes (e.g., BD and soil structure information). The use of globally-derived PTFs embedded in various models, on the other hand, resulted in larger error in SWRC estimation which in turn ends up with large deviation in model outcomes as compared to those based on measured data. To sum up, cautious functional evaluation of PTFs should be done before implementing them as alternative data sources for modeling, as 'blind' application can result in wrong decisions in sustainable natural resources management.

7.4. Conclusions

The functional evaluation has shown that the performance of agro-hydrological models is sensitive to the variation in soil hydraulic parameters. In three of the four investigated soils, locally derived PTFs for VMD soils offered more accurate estimation of soil hydraulic properties, and as a result of that, they also performed well in simulating different functional hydraulic responses of paddy soils in wet and dry growing seasons using both a conceptual crop model and a physically-based hydrological model. The explanation of these results could probably be the site-specific feature and pedo-agro-climatological dependency of PTFs which made them more representative and applicable to soils in the region for which the PTFs have been determined. Indeed, it is widely known that the performance of a PTF may vary with pedological and geomorphological origin of the soil on which it was developed (Minasny et al., 1999), hence the application of globally-offered PTF's may be a source of error in model simulation. This study strengthened the warning raised by Leenhardt (1995) that the user of a soil water model must be concerned not only with the reliability of the model itself and its sensitive to changes in parameters, but also with the reliability of the facilities that could be made available.

It is also important to highlight the worse performance of indirect estimation methods with PTFs in the case of salt-affected soils in Soc Trang province. Such findings are explainable due to the fact that dynamic hydraulic behaviors of saline-soils could not be properly captured or accurately predicted by the empirical functions derived from either international databases or local data sets. Indeed, swelling and/or dispersion of soil colloids by salts (primarily Na) alters the geometry of pore systems and make them very dynamic, hence their corresponding SWRC cannot effectively be described from static soil characteristics such as texture and soil pore distribution. Moreover, as the local data was collected based on cross-sectional sampling, it is limited in capturing the temporal change in hydraulic behavior of soil which is affected by salinity from seasonal tidal movement.

To sum up, the applications of PTFs to soils which are different from or not represented by the data used to develop regression functions should be done with caution, particularly in hydrological model applications, as their error in soil hydraulic properties estimation can jeopardize the modeling results of soil water balance components which in turn can affect management and planning guidance for natural resources management and agricultural production (e.g., improper irrigation scheduling, or wrong forecasting of ground water contamination). Well-tested PTFs, on the other hand, can be considered as a reasonable, reliable and cheap alternative to sampling and laboratory measurements of soil hydraulic data for modeling, especially in large-scale research projects. The improvement of the predictive capacity of locally-derived PTFs can be made by increasing their coverage in terms of temporal and spatial variability of the soils.

This evaluation study helps us to increase our confidence in the usefulness of locally derived PTFs in hydrological modeling. However, in practice, it is known that the error derived from estimation of soil hydraulic properties is seldom the only one arising from simulation of the soil water balance (as reported in the studies of Espino et al.,1996; Leenhardt, 1995; Minasny and McBratney, 2002; Vereecken et al., 1992). With all the uncertainties involved, uncertainty analysis should be extended to identify the contribution as well as the interaction of individual sources of errors to the overall uncertainty of model outcomes.

Chapter 8

GENERAL CONCLUSIONS AND RECOMMENDATIONS

As was clearly stated by Lin (2012), soil and water are two critical natural resources that are fundamentally supports to ecosystem services and human well-being. Indeed, the interaction between soil and water builds up the vital interface between the abiotic and biotic worlds, and supports life of all kinds on Earth. However, our understanding of the complexity of soil and water interactions in the landscape remains remarkably limited (Pachepsky et al., 2015), particularly in tropical regions, where the largest population growth occurs and the problems of food insecurity, soil degradation, climate change and water scarcity are greatest (Lal, 2000; Minasny and Hartemink, 2011).

Pedotransfer functions emerged as relationships between soil hydraulic parameters and the easier measurable soil properties usually available from soil survey (Pachepsky et al., 2006). PTFs currently serve as an essential tools for diagnostics, monitoring, predictions, and management of soil and water resources as a life-supporting Earth system. Developing and improving pedotranfer functions is known as the mainstream way of packaging data and knowledge of soil hydraulic functioning for a wide-range of hydrological, agricultural, and environmental modelling applications in large scale research projects or in poor-data environments such as the Vietnamese Mekong Delta.

The main objective of the research in this dissertation was to develop innovative SWRC-PTFs for tropical VMD soils in order to better understand the complex soil-water relationships in the studied region where soils have been mainly exploited for paddy rice production. Such soil hydraulic information is prerequisite for sustainable soil-water management in VMD.

In this chapter, we briefly summarize our main findings and the contributions of this research to the achievement of the study’s objectives. General conclusions are organized in three parts: (1) the applicability of published PTFs for VMD soils, (2) the important strategies in PTF development for the soils of interest, and (3) the utility of the derived PTFs in agro-hydrological modeling. Afterwards, some recommendations for future researches are presented.

8.1. The applicability of SWRC-PTFs for soils in VMD

Since PTFs are empirical relationships calibrated based on information of training data, the predictive performance of particular PTFs is strongly dependent on the quality of the training data in terms of their coverage and representativeness to the target soils that need SWRC prediction. As it is commonly reported in PTF evaluation studies (e.g., Botula et al. (2012); Cornelis et al. (2001); Nebel et al. (2010); Nemes et al. (2003)), the predictive performance of published PTFs varies with the pedo-agro-climatic origin of the soils from which they were developed. Extrapolation of PTFs beyond the geographical locations and the statistical limits of the calibration data sets should be avoided or at least carefully evaluated for their predictive quality before use (Cornelis et al., 2001).

In **Chapter 3**, several well-known statistical regression and pattern-recognition PTFs which were developed for soils in both tropical and temperate climates were evaluated in terms of their applicability and reliability in predicting SWRC of VMD soils. By assessing the correspondence between measured and PTF predicted values of SWRC at field capacity (FC) and permanent wilting point (PWP), the evaluation showed that the predictive performance of published PTFs for tropical VMD soils greatly varied among used PTFs and that it was more dependent on the coverage and the quality of PTFs' calibrated databases. Specifically, the prediction errors by using 'tropical' PTFs were mainly attributed to the difference in clay mineralogy, while unreliable predictions of 'temperate' PTFs might result from the discrepancy in the distribution of soil texture classes between training databases and testing data set. The statistical regression-based PTFs of Adhikary et al. (2008); Botula (2013); Minasny and Hartemink (2011), and the pattern-recognition 'k-Nearest' PTFs which were derived from soils in tropical regions performed more reliable to VMD soils than those developed for soils in temperate climates. There were also exceptional 'temperate' PTFs which performed rather well, i.e. the regression-based PTFs of Saxton and Rawls (2006) and the ANN-based "Rosetta" PTFs of Schaap et al. (2001).

Generally, PTFs derived from large databases of soils in the regions having similar climatological and pedological conditions to VMD soils performed more reliable than others. The applicability index, which was determined based on the validity range of input variables used in regression equations, displayed a significant correlation with the predictive error (expressed by RMSE) of the applied PTFs. Such index, together with the geographical information of published PTFs, can be used as integral indicator to select appropriate PTFs in cases no SWRC data and no specific PTFs are available for timely uses.

The findings in **Chapter 3** suggested that estimates of soil water retention by the investigated 'temperate' and 'tropical' PTFs might induce errors which probably prolong to the outputs of agro-hydrological models that might be used in various agricultural studies in VMD (see section 8.3) or other regions having similar pedo-geo-climatic conditions. Large discrepancy in the derived soil hydraulic data can substantially reduce the quality of the modelling results, hence, specific SWRC-PTFs for tropical delta soils exploited for paddy rice production need to be developed.

In the next section, two important PTF development strategies, i.e. determination of significant predictors of SWRC estimation and of regression methods used to derive predictive models, were investigated in order to obtain the most accurate and reliable VMD-PTFs.

8.2. Important strategies in PTFs development

8.2.1. Potential predictors of SWRC-PTFs

As it was clearly stated by Minasny and Hartemink (2011), developing PTFs should not be a statistical exercise, but rather a physical basis to select the appropriate predictors for predicting a variable. These authors, among others (e.g. Botula et al.2013; Khlosi et al., 2013; Vereecken et al., 2010), have stressed that PTFs developers should understand the soils and use this knowledge to select logic predictors. Identification of additional soil information that could improve the accuracy of the PTFs, besides classical PTF predictors, is hence one of the key areas in PTF's research.

Regarding SWRC's predictor variables, particle size distribution, OM or OC content, and BD have been considered as the most basic and common predictors used in SWRC-PTFs (Botula et al., 2013; Minasny and Hartemink, 2011; Saxton and Rawls, 2006). Soil structure (defined as the spatial arrangement of soil particle in secondary units) has been shown to affect soil hydraulic properties (Pachepsky et al. 2003). Typical PTF predictors, such as soil texture, BD and OM content are related to soil structure in a broad sense, but are not sufficient to define the spatial arrangement of structural units which is a key factor affecting the ability of soil to retain and transmit water. Therefore, including soil structure information in textured-based PTFs is expected to improve the certainty of PTF prediction (Pachepsky and Rawls, 2003). Although several authors (Pachepsky et al., 2006; Vereecken et al.,

2010) pointed out the importance of using soil structure in PTF development, very few studies really tried to account for it.

In **Chapter 4**, basic soil properties (i.e. soil texture, soil OM, and BD) together with other available soil information in our regional data set (e.g. descriptive soil structure information, pH, EC, plastic limit, stability index) were considered in identifying significant predictors variables for SWRC-PTFs in VMD. A classical stepwise multiple linear regression technique which enables to automatically detect significant input variables in regression equations was utilized in this study. The results revealed that SWRC of tropical VMD soils could be satisfactorily estimated by classical PTF predictors (e.g., sand, silt, clay content, BD, and OC). Moreover, incorporating descriptive soil structure information (e.g., presence or absence of pedality) as grouping criterion priori to PTF development did improve the prediction accuracy of SWRC, especially in the wet moisture range. Plastic limit was found to be a promising predictors for SWRC-PTFs of soils having a given degree of structural development.

The use of a simple categorical presentation of soil structure in developing PTFs matches well the two basic principles in PTF research, i.e. (1) the principle of efficiency, and (2) the principle of uncertainty. Indeed, first, descriptive soil structural information is available in most soil survey databases; its readiness for use in PTF development is substantially high. Second, descriptive soil structural information was shown to have an influence on soil hydraulic properties in the macro-pore flow zone (Pulido Moncada et al., 2014). Including soil structure information in texture-based PTFs is thus expected to improve the certainty of PTF predictions, particularly in the wet and intermediate range of SWRC.

8.2.2. Regression methods

Beside specifying the significant predictors for regional VMD-PTFs, determination of appropriate regression methods is also important to obtain best-performing PTFs. The regression methods commonly used for SWRC-PTFs development are generally grouped into two main categories: statistical regression techniques (Vereecken and Herbst, 2004), and data mining or pattern-recognition techniques (Pachepsky and Schaap, 2004; Vereecken et al., 2010). Although statistical regression methods (e.g. MLR) offer simple, reasonable and well-interpretable models, they also revealed several drawbacks. Specifically, with the emergence of large soil databases in the tropical regions nowadays (Botula et al., 2014; Minasny and Hartemink, 2011), classical statistical regression techniques (e.g. MLR) are limited in detecting important relations in such vast data space. Data-mining or pattern-recognition techniques that are flexible enough to handle large data are increasingly needed. These techniques are theoretically able to extract the most important information, and to uncover previously unknown patterns of the relationships between soil properties in the database that may be hidden from MLR (Botula et al., 2013).

To obtain conclusive results in PTFs studies, large soil hydraulic databases of good quality are usually required. However, the lack of well-defined hydraulic databases or the availability of rather limited data sets is the practical reality in many developing countries located in the tropics (e.g., VMD). The challenge, however, is to cope with these limitations when developing specific PTFs for soils in these regions. In this study (**Chapter 5** and **Chapter 6**), we tried to evaluate the potential of different data mining techniques to predict SWRC when taking significant soil predictors determined in **Chapter 4**.

The predictive capabilities of point PTFs and pseudo-continuous (PC) PTFs developed by Multiple Linear Regression (MLR), Artificial Neural Networks (ANN), Support Vector Machine for Regression (SVR), and k-Nearest Neighbors (kNN) methods were compared and evaluated in **Chapter 5**. The results showed that point PTFs derived by various regression techniques provided comparable accuracy in estimating SWRC at specific matric potentials, but the reliability of ANN, SVR and kNN models was much better than that of MLR. In case of PC-PTFs, ANN and kNN models outperformed SVR and MLR in both training and testing phases. Our findings confirm the superiority of data-mining approaches in modeling the complex system of soil and water even when a limited dataset is available (average RMSE for the test data set varied from 0.049 to 0.053 m^3 m^{-3}). Although the non-parametric kNN method has a constraint in estimating SWRC in a pseudo-continuous manner, this method has great benefits due to its flexibility, simplicity, accuracy and capacity to append new observations without redeveloping the PTFs.

Using pseudo-continuous PTFs has the advantage that we need only one predictive function for the continuous estimation of the whole SWRC. The results of this study (**Chapter 5**) and those of Haghverdi et al. (2012); Haghverdi et al. (2014), showed that the ANN technique acts as a unique method for developing PC-PTFs. Others, e.g., MLR and SVM, seem do not appropriate. Theoretically, the non-parametric kNN technique is limited in representing the PC-PTFs' philosophy (i.e. predicting SWRC in a continuous manner without using any fitted SWRC equations).

The PTF predictive performance might be even more enhanced by the combined effect of the two above-mentioned strategies, i.e., adding significant predictors as descriptive soil structural information (**Chapter 4**), and implementing more flexible regression algorithms (**Chapter 5**). The main objective of **Chapter 6** was to investigate whether the improved effects of categorical soil structure information found in Chapter 4 with MLR could be enduringly captured by the best-performing SVM and kNN techniques (Chapter 5) for point estimation. The results of the study showed that incorporating descriptive soil structure information improved the accuracy of PTFs derived by the SVM approach in the range of matric potentials of -6 to -33 kPa (average RMSE decreased up to 0.005 m^3 m^{-3} after grouping, depending on matric potentials). The improvement was primarily attributed to the outperformance of SVM-PTFs calibrated on data of the structureless soil group. No improvement was obtained with the kNN technique, at least not in our study in which the data set became limited in size after grouping.

The use of pedological soil structure information may also have drawbacks because (1) soil structure is described in qualitative rather than quantitative terms, hence it is impractical to directly use this information as SWRC predictor; and (2) structure characterization is usually done at a scale that is too coarse to reveal the arrangement of the fine pores that retain water at low soil matric potentials (i.e. micro-pore flows). Anyhow, the attempts to use soil structural information in the water retention PTFs have shown some improvement in their accuracy as manifested in Chapter 4 and Chapter 6 as it was suggested in the studies of William et al. (1992), Rawls and Pachepsky (2002), Pachepsky et al. (2006).

This study was an attempt to see how descriptive soil structural information can potentially improve the prediction accuracy of PTFs and whether it might be an important strategy to include it, while considering different regression techniques. Since there is an impact of the regression techniques on the improved effect of incorporating qualitative soil structure information, selecting a proper technique will help to maximize the combined influence of flexible regression algorithms and soil structure information on PTF accuracy.

8.3. PTFs' utility for practical agro-hydrological modeling

Once the PTFs were constructed, their utility was assessed by using functional validation. Indeed, since the PTFs is not an aim on itself, it is better to evaluate the capacity of the derived PTFs on specific applications. Modeling water and solute transport has become an important tool in simulating agricultural productivity as well as environmental quality. The use of agro-hydrological models, however, is often limited by the lack of information on soil hydraulic properties. For many applications, using approximation of the hydraulic parameters offered by PTFs embedded in the models is an attractive alternative of missing data (Nemes et al., 2006a). In **Chapter 7**, a functional evaluation was conducted in which the effect of replacing direct laboratory measured SWRC by PTF-predicted values (obtained from locally derived VMD-PTFs and globally-offered PTFs) on the outcomes of agro-hydrological models (e.g., crop-water model AquaCrop and hydrological model Hydrus-1D in this study) was investigated.

The performance of agro-hydrological models is known to be sensitive to the variation in soil hydraulic parameters (Espino et al., 1996; Georgoussis et al., 2009; Nemes et al., 2010). The functional evaluation in Chapter 7 has shown that for soils that are typically found in the VMD region (e.g. alluvial soil with clayey textured, acid sulfate soils with high OM content in the surface, and sandy soils), locally derived VMD-PTFs offer more accurate estimation of SWRC, and consequently also perform rather well (as compared with the simulation scenario using laboratory-measured SWRC data) in simulating different hydraulic responses of paddy fields in wet and dry growing seasons using both a conceptual crop model and a physically-based hydrological model. The explanation of these results could probably be the site-specific feature and pedo-agro-climatological dependency of the PTFs which made that they are more representative and applicable to soils in the region for which the PTFs were developed.

On the other hand, the results in **Chapter 7** also raised awareness towards the arbitrarily application of globally-offered PTF's in agro-hydrological models (e.g. PTFs of Saxton and Rawls, 2006, used in "Soil Water Characteristics" software and

ANN-PTFs of Schaap et al.,2001 in “Rosetta” software) which may thus be a source of error in the model’s simulation. The statistical and functional evaluations of published PTFs in this study (**Chapter 3** and **Chapter 7**) again strengthened the warning raised by Leenhardt (1995) that the users of a soil water model must be concerned not only with the reliability of the model itself and its sensitive to changes in parameters, but also with the reliability of the facilities that could be made available.

It’s also important to highlight the poor performance of all PTFs in the case of salt-affected soil in the VMD. This could be explained by the fact that dynamic hydraulic behavior of salt-affected soils could not be properly captured or accurately predicted by the empirical functions of basic soil properties derived based on either international databases or the local data set. Indeed, swelling and/or dispersion of soil colloids by salts alters the geometry of pore systems and made them very dynamic. Capturing the temporal change in hydraulic behavior of salt-affected soils requires further research which will be discussed later in section 8.4.

To sum up, the applications of PTFs to soils which are different from or not represented by the data used to calibrate regression functions should be done with caution, particularly in hydrological model applications as their error in soil hydraulic properties estimation can jeopardize the modeling results of soil water balance components which in turn can affect management and planning guidance for natural resources management and agricultural production. Well-tested PTFs on the other hand, can be considered as a reasonable, reliable and cheap alternative to sampling and laboratory measurements of soil hydraulic data for modeling, especially in large-scale research projects. The improvement of the predictive capacity of locally-derived PTFs can be made by increasing their coverage in terms of temporal and spatial variability of the soils.

8.4. Recommendations for future research

This study attempted to develop innovative SWRC-PTFs for the tropical Vietnamese Mekong Delta. While studying this, several issues which require further clarification were revealed. Following specific aspects need further investigation:

- The presence of indicators that help to select appropriate SWRC-PTFs available from literature for specific regional soil-water management applications are needed. The validity of the proposed integral indicator in **Chapter 3** (i.e.. applicability index together with the information of geographical domain of regions where the PTFs were derived) need further confirmation by other studies with different investigated soils and published PTFs.

- There is a concern about the limitation of using rather small data sets to derive and validate conclusive PTFs, particularly those derived by highly data demanding pattern-recognition techniques. More research on finding thresholds for data set sizes is recommended as this is yet unknown.

- In light of the findings of this dissertation, in that including pedological soil structure description might lead to better predictions, but anyhow to a limited extent, the use of a (semi)quantitative approach of soil structure evaluation might be an alternative and perhaps better option, or at least worth for further investigations. Visual Soil Assessment (VSA) approaches similarly use morphological features that are described in the field when giving scores to soil structure. The classical soil structure description obtained from soil survey presents structural information under categories or classes, while VSA presents the outcome of visual assessment by a numeric score-system (in an agronomical suitability perspective). Using scores might be a more flexible approach for PTF development since scores can be used both as grouping criteria in statistical regression techniques and as a direct predictor of SWRC in both statistical regression and data-mining approaches. VSA is highlighted here as a prospective alternative of soil structure description for future research when incorporating soil structure in PTFs. VSA approaches are now receiving great interest from soil scientist and agronomists; such information is expected to become available in many soil databases in the near future, hence warrant their potential for implementing in PTF development and utilization.

- Ignoring the temporal dynamics of soil hydraulic properties in PTF development might create several hurdles in practical soil water management, particularly for salt-affected soil as experienced in **Chapter 7**. Indeed, Pachepsky et al. (2015) recently argued that a temporal component might be required in PTF development to improve their reliability. Longitudinal study design for data collection (e.g., combining multiple sample timings) which may help to overcome temporal variability, hence, need future research for verification.

- Further improvement of PTFs predictability for salt-affected soils is highly recommended since the area of such specific soils has been increased nowadays in tropical lowland deltas due to the effect of climate change (Mondal et al., 2015). Beside incorporating a temporal component in data collection (as mention above), further investigations about additional significant soil variables (e.g., EC, SAR, to mention a few) to explain the variability in SWRC behaviour of salt-affected soils need to be carried out.

- Together with soil water retention characteristics, soil hydraulic conductivity functions (HCF) are also key ingredients in most analytical and numerical models for flow and transport in unsaturated porous media. However, the reliable data of hydraulic conductivity functions seems even more difficult to obtain than SWRC due to its high variability in both time and space and its scale-dependency. The question about which methods (both in-situ field measurement and ex-situ laboratory analysis) could be considered as the standard one for characterizing soil hydraulic conductivity functions is still controvertible. Hence, further research on soil hydraulic conductivity is needed to define robust methods to determine or predict them in a reliable way based on easily measurable and readily available soil properties.

REFERENCES

Abbaspour, K.C., Moon, D.E., 1992. Relationships between conventional field information and some soil properties measured in the laboratory. Geoderma. 55(1–2), 119-140.

Adhikary, P.P. et al., 2008. Pedotransfer functions for predicting the hydraulic properties of Indian soils. Australian Journal of Soil Research. 46(5), 476-484.

Aimrun, W., Amin, M.S.M., 2009. Pedo-transfer function for saturated hydraulic conductivity of lowland paddy soils. Paddy Water Environment. 7(3), 217-225.

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(1), 55-58.

Alexander, I., Morton, H., 1990. An Introduction to Neural Computing. Chapman and Hall, London.

Ali, M.H., Biswas, T.D., 1986. Soil water retention and release as related to mineralogy of soil clays, Proceedings of 55th Indian Science Congress 3, 633.

Allen, R.G., Pereira, L.S., Raes, D., Smith, M., 1998. Crop evapotranspiration - Guideline for computing crop water requirements - FAO Irrigation and Drainage paper 56, FAO, Rome, Italy.

Anderson, J.L., Bouma, J., 1973. Relationships Between Saturated Hydraulic Conductivity and Morphometric Data of an Argillic Horizon1. Soil Science Society of American Journal. 37(3).

Andrade, A.R.S., Paixao, F.J.R., Azevedo, C.A.V., Neto, J.D., 2008. The soil available water estimate through the fractal theory. Pesquisa Aplicada & Agrotecnologia. 1, 157-162.

Araya, A., Keesstra, S.D., Stroosnijder, L., 2010. Simulating yield response to water of Teff (Eragrostis tef) with FAO's AquaCrop model. Field Crops Research. 116(1–2), 196-204.

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 American Journal. 45(6), 1023-1030.

ASTM, 2010. Standard test methods for liquid limit, plastic limit and plasticity index of soils. ASTM International, West Conshohocken, PA.

Azooz, R.H., Arshad, M.A., Franzluebbers, A.J., 1996. Pore Size Distribution and Hydraulic Conductivity Affected by Tillage in Northwestern Canada. Soil Science Society of American Journal. 60(4).

Babaeian, E., Homaee, M., Montzka, C., Vereecken, H., Norouzi, A.A., 2015. Towards Retrieving Soil Hydraulic Properties by Hyperspectral Remote Sensing. Vadose Zone Journal. 14(3).

Bannayan, M., Hoogenboom, G., 2009. Using pattern recognition for estimating cultivar coefficients of a crop simulation model. Field Crops Research. 111(3), 290-302.

Bastiaanssen, W.G.M., Allen, R.G., Droogers, P., D'Urso, G., Steduto, P., 2007. Twenty-five years modeling irrigated and drained soils: State of the art. Agricultural Water Management. 92(3), 111-125.

Batjes, N.H., 1996. Development of a world data set of soil water retention properties using pedotransfer rules. Geoderma. 71(1–2), 31-52.

Bayat, H., Neyshaburi, M.R., Mohammadi, K., Nariman-Zadeh, N., Irannejad, M., 2013. Improving water content estimations using penetration resistance and principal component analysis. Soil and Tillage Research. 129, 83-92.

Baydaroğlu, Ö., Koçak, K., 2014. SVR-based prediction of evaporation combined with chaotic approach. Journal of Hydrology. 508(0), 356-363.

Bell, M.A., van Keulen, H., 1995. Soil pedotransfer functions for four Mexican soils. Soil Sci Soc Am J. 59.

Bhattacharya, T., Panda, R.K., 2013. Effect of climate change on rice yield at Kharagpur, West Bengal. IOSR Journal of Agriculture and Veterinary Science. 4(2), 6-12.

Børgesen, C.D., Iversen, B.V., Jacobsen, O.H., Schaap, M.G., 2008. Pedotransfer functions estimating soil hydraulic properties using different soil parameters. Hydrological Processes. 22(11), 1630-1639.

Børgesen, C.D., Schaap, M.G., 2005. Point and parameter pedotransfer functions for water retention predictions for Danish soils. Geoderma. 127(1–2), 154-167.

Botula, Y.-D., Van Ranst, E., Cornelis, W.M., 2014. Pedotransfer functions to predict water retention for soils of the humid tropics: a review. Revista Brasileira de Ciência do Solo. 38, 679-698.

Botula, Y.D., 2013. Indirect methods to predict hydrophysical properties of soils of Lower Congo. PhD thesis Thesis, Ghent University, Gent, Belgium, 236 pp.

Botula, Y.D., Cornelis, W.M., Baert, G., Van Ranst, E., 2012. Evaluation of pedotransfer functions for predicting water retention of soils in Lower Congo (D.R. Congo). Agricultural Water Management. 111, 1-10.

Botula, Y.D., Nemes, A., Mafuka, P., Van Ranst, E., Cornelis, W.M., 2013. Prediction of water retention of soils from the humid tropics by the nonparametric k-nearest neighbor approach. Vadose Zone Journal. 12(2), 1-17.

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 American Journal. 63(6).

Brooks, R.H., Corey, A.T., 1964. Hydraulic properties of porous media, Hydrology Paper 3. Colorado State University, Fort Collins, Colorado, USA.

Bruand, A., 2004a. Preliminary grouping of soils. In: Pachepsky, Y., Rawls, W.J. (Eds.), Developments in Soil Science. Elsevier, pp. 159-174.

Bruand, A., 2004b. Utilizing mineralogical and chemical information in PTFs. In: Pachepsky, Y., Rawls, W.J. (Eds.), Developments in Soil Science. Elsevier, pp. 153-158.

Buccigrossi, F., Caliandro, A., Rubino, P., Mastro, M.A., 2010. Testing some pedo-transfer 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. 5(4), 367-382.

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(11), 2761-2776.

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 American Journal. 65(3), 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.

Chakraborty, D. et al., 2011. Pedotransfer functions for predicting points on the moisture retention curve of Indian soils. Indian Journal of Agricultural Sciences. 81, 1030-1036.

Christiaens, K., Feyen, J., 2001. Analysis of uncertainties associated with different methods to determine soil hydraulic properties and their propagation in the distributed hydrological MIKE SHE model. Journal of Hydrology. 246(1–4), 63-81.

Clapp, R.B., Hornberger, G.M., 1978. Empirical equations for some soil hydraulic properties. Water Resources Research. 14(4), 601-604.

Coen, G.M., Wang, C., 1989. Estimating vertical saturated hydraulic conductivity from soil morphology in Alberta. Canadian Journal of Soil Science. 69(1), 1-16.

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 American Journal. 69(6), 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 American Journal. 65(3), 638-648.

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(6), 682-690.

Danalatos, N.G., Kosmas, C.S., Driessen, P.M., Yassoglou, N., 1994. Estimation of the draining soil moisture characteristic from standard data as recorded in routine soil surveys. Geoderma. 64(1–2), 155-165.

Dane, J.H., Topp, G.C., 2002. Methods of soil analysis. Part 4. Physical methods. Soil Science Society of America Book series. Soil Science Society of America, Madison, Wisconsin, USA.

Dasarathy, B.V., 1991. Nearest neighbor (NN) norms: NN pattern classification techniques. IEEE Computer Society Press.

Dash, C.J., Sarangi, A., Singh, D.K., Singh, A.K., Adhikary, P.P., 2015. Prediction of root zone water and nitrogen balance in an irrigated rice field using a simulation model. Paddy Water Environment. 13(3), 281-290.

Dashtaki, S.G., Homaee, 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(1), 68-74.

De Leenheer, L., De Boodt, M., 1959. Determination of aggregate stability by the change in mean weight diameter, Proceedings of International Symposium of Soil Structure. . Mededelingen van de Landbouwhogeschool: Gent, Ghent, Belgium, pp. 290-300.

Demuth, H., Beale, M., 2000. Neural Network Toolbox. Mathworks, Inc.

Dexter, A.R., Bird, N.R.A., 2001. Methods for predicting the optimum and the range of soil water contents for tillage based on the water retention curve. Soil and Tillage Research. 57(4), 203-212.

Dijkerman, J.C., 1988. An Ustult-Aquult-Tropept catena in Sierra Leone, West Africa, II. Land qualities and land evaluation. Geoderma. 42(1), 29-49.

Doorenbos, J., Kassam, A.H., 1979. Yield response to water Irrigation and Drainage Paper No. 33. FAO, Rome, Italy, 193 pp.

Driessen, P., Dudal, R., 1989. Geography formation, properties and use of the major soils of the world. The Netherlands and Catholic University, Luven.

Efron, B., Tibshirani, R.J., 1993. An introduction to the bootstrap. Monographs on Statistics and Applied Probability. Chapman and Hall, New York.

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(10), 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(10), 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(3), 235-253.

Estellès, P., Jensen, H., Sánchez, L., Vechiu, G., 2002. Sustainable Development in the Mekong Delta. Fysisk Institut, Center for Environmental Studies, University of Aarhus.

FAO, 1974. Management properties of Ferrasols, FAO Soils Bulletin 23. Food and Agriculture Organization of the United Nations, Rome, Italy.

FAO, 2006. Soil description. In: Jahn, R., Blume, H.P., Asio, V.B., Spaargaren, O., Schad, P. (Eds.), Guideline for soil description 2006. FAO, Rome, Italy, pp. 21-64.

FAO, 2009. ETo Calculator, Land and Water Digital Media Series No. 36. FAO, Rome, Italy.

FAO, 2015. Healthy soils are the basis for healthy food production. FAO, Rome, Italy.

FAOSTAT, 2012. Food and Agriculture Organization of the United Nations, Rome, Italy.

Farahani, H.J., Izzi, G., Oweis, T.Y., 2009. Parameterization and Evaluation of the AquaCrop Model for Full and Deficit Irrigated Cotton Agronomy Journal. 101(3).

Feddes, R.A., Kowalik, P.J., Zaradny, H., 1978. Simulation of Field Water Use and Crop Yield. John Wiley, New York, NY.

Feder, J., 1988. Fractals. Plenum Press, New York, 283 pp.

Gaiser, T., Graef, F., Cordeiro, J., eacute, Carvalho, 2000. Water retention characteristics of soils with contrasting clay mineral composition in semi-arid tropical regions. Soil Research. 38(3), 523-536.

García-Vila, M., Fereres, E., Mateos, L., Orgaz, F., Steduto, P., 2009. Deficit Irrigation Optimization of Cotton with AquaCrop. Agronomy Journal. 101(3).

Gärdenäs, A.I., Hopmans, J.W., Hanson, B.R., Šimůnek, J., 2005. Two-dimensional modeling of nitrate leaching for various fertigation scenarios under micro-irrigation. Agricultural Water Management. 74(3), 219-242.

Gee, G.W., Bauder, J.W., 1986. Particle-size analysis. In: Klute, A. (Ed.), Methods of soil analysis. Part 1. Physical and Mineralogical methods. American Society of Agronomy, Madison, Wisconsin, USA, pp. 383-411.

Geerts, S. et al., 2009a. Simulating Yield Response of Quinoa to Water Availability with AquaCrop. Agronomy Journal 101(3), 499-508.

Geerts, S. et al., 2009b. Modeling the potential for closing quinoa yield gaps under varying water availability in the Bolivian Altiplano. Agricultural Water Management. 96(11), 1652-1658.

Georgoussis, H., Moussouliotis, A., Babajimopoulos, C., 2009. Functional evaluation of pedotransfer functions in irrigation scheduling. In: Topping, B.H.V., Tsompanakis, Y. (Eds.), the First International Conference on Soft Computing Technology in Civil, Structural, and Environmental Engineering. Civil-Comp Press, Stirlingshire, Scotland.

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(1–2), 77-106.

Giménez, 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(2), 79- 87.

Grossman, R.B., Reinsch, T.G., 2002. Bulk Density and Linear Extensibility. In: Dane, J.H., Topp, G.C. (Eds.), Methods of Soil Analysis: Physical Methods, Part 4. Soil Science Society of America, Madison, USA, pp. 201-228.

GSO, 2009. International Merchandise Trade Vietnam 2009. General Statistics Office, Ha Noi, Vietnam.

Guber, A., Pachepsky, Y., Shein, E., Rawls, W.J., 2004. Soil aggregates and water retention. In: Pachepsky, Y., Rawls, W.J. (Eds.), Developments in Soil Science. Elsevier, pp. 143-151.

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(6), 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. 442–443, 46-54.

Haghverdi, A., Leib, B.G., Cornelis, W.M., 2015. A simple nearest-neighbor technique to predict the soil water retention curve. Transactions of the ASABE. 58(3), 697-705.

Haghverdi, A., Öztürk, H.S., Cornelis, W.M., 2014. Revisiting the pseudo continuous pedotransfer function concept: Impact of data quality and data mining method. Geoderma. 226–227, 31-38.

Hansen, V.E., Israelsen, O.W., Stringham, G.E., 1980. Irrigation principles and practices, 4th edition. Wiley, NewYork, 417 pp.

Hastie, T., Tibshirani, R., Friedman, J., 2009. The elements of statistical learning: data mining, inference, and prediction. Springer Series in Statistics. Springer New York.

Haverkamp, R., Parlange, J.Y., 1986. Predicting the water-retention curve from particle-size distribution: 1. Sandy soils without organic matter. Soil Science. 142(6), 325-339.

Heng, L.K., Hsiao, T., Evett, S., Howell, T., Steduto, P., 2009. Validating the FAO AquaCrop Model for Irrigated and Water Deficient Field Maize Agronomy Journal. 101(3).

Hillel, D., 1980. Fundamentals of soil physics. Academic Press, New York.

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 functions developed for tropical soils. Geoderma. 108(3-4), 155-180.

Hong, W.-C., Pai, P.-F., 2007. Potential assessment of the support vector regression technique in rainfall forecasting. Water Resource Management. 21(2), 495-513.

Horn, R., Baumgartl, T., 1999. Dynamic processes in structured soils. In: Sumner, M., al., e. (Eds.), Handbook of Soil Science. CRC Press, Boca Raton, FL, USA, pp. A19-A46.

Hsiao, T.C. et al., 2009. AquaCrop—The FAO Crop Model to Simulate Yield Response to Water: III. Parameterization and Testing for Maize. Agronomy Journal. 101(3).

Hu, J., Liu, J., Liu, Y., Gao, C., 2011. EMD-KNN Model for Annual Average Rainfall Forecasting. Journal of Hydrologic Engineering. 18(11), 1450-1457.

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(1–4), 325-334.

IBM, 2011. SPSS Statistics for Window. Version 20.0. IBM Corp., Armonk, New York, USA.

IPCC, 2007. Climate change 2007 Working Group II: Impacts, Adaptation and Vulnerability, Cambridge, United Kingdoms and New York, NY, USA.

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(5), 1437-1444.

Jalota, S.K. et al., 2012. Mitigating future climate change effects by shifting planting dates of crops in rice–wheat cropping system. Regional Environmental Change. 12(4), 913-922.

Jana, R.B., Mohanty, B.P., 2011. Enhancing PTFs with remotely sensed data for multi-scale soil water retention estimation. Journal of Hydrology. 399(3-4), 201-211.

Janik, L.J., Merry, R.H., Forrester, S.T., Lanyon, D.M., Rawson, A., 2007. Rapid Prediction of Soil Water Retention using Mid Infrared Spectroscopy. Soil Science Society of American Journal. 71(2).

Jin, X.-l. et al., 2014. Assessment of the AquaCrop Model for Use in Simulation of Irrigated Winter Wheat Canopy Cover, Biomass, and Grain Yield in the North China Plain. PLoS ONE. 9(1), e86938.

Jong, R.D., McKeague, J.A., 1987. A comparison of measured and modelled soil water retention data. Canadian Journal of Soil Science. 67(3), 697-703.

Katambara, Z. et al., 2013. Optimizing system of rice intensification parameters using AquaCrop model for increasing water productivity and water use efficiency on rice production in Tanzania. Journal of Agriculture and Sustainability. 4(2), 235-244.

Kay, B.D., Angers, D.A., 2002. Soil Structure. In: Warrick, A.W. (Ed.), Soil physics companion. CRC Press, Boca Raton, FL, pp. 249-295.

Keller, T., Dexter, A.R., 2012. Plastic limits of agricultural soils as functions of soil texture and organic matter content. Soil Research. 50(1), 7-17.

Kern, J.S., 1995. Evaluation of Soil Water Retention Models Based on Basic Soil Physical Properties. Soil Science Society of American Journal. 59(4), 1134-1141.

Khlosi, M., 2003. Physical-conceptual models to predict the soil moisture retention curve. Master Dissertation Thesis, Ghent University, Ghent, Belgium.

Khlosi, M., Alhamdoosh, M., Douaik, A., Cornelis, W.M., 2016. Enhanced pedotransfer functions with support vector machines for predicting water retention of calcareous soil. European Journal of Soil Science. (in press).

Khlosi, M. et al., 2013. Exploration of the Interaction between Hydraulic and Physicochemical Properties of Syrian Soils. Vadose Zone Journal. 12(4), 1-11.

King, J.J., Franzmeier, D.P., 1981. Estimation of Saturated Hydraulic Conductivity from Soil Morphological and Genetic Information. Soil Science Society of American Journal. 45(6), 1153-1156.

Klute, A., 1982. Tillage Effects on the Hydraulic Properties of Soil: A Review. In: Unger, P.W., Van Doren Jr, D.M. (Eds.), Predicting Tillage Effects on Soil Physical Properties and Processes. ASA Special Publication. American Society of Agronomy and Soil Science Society of America, Madison, WI.

Kodaira, M., Shibusawa, S., 2013. Using a mobile real-time soil visible-near infrared sensor for high resolution soil property mapping. Geoderma. 199, 64-79.

Koekkoek, E.J.W., Booltink, H., 1999. Neural network models to predict soil water retention. European Journal of Soil Science. 50(3), 489-495.

Kögel-Knabner, I. et al., 2010. Biogeochemistry of paddy soils. Geoderma. 157(1–2), 1-14.

Kutner, M.H., Nachtsheim, C.J., Neter, J., Li, W., 2005. Applied linear statistical models, 6. McGraw-Hill New York, United States.

Lal, R., 1978. Physical properties and moisture retention characteristics of some nigerian soils. Geoderma. 21(3), 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.

Lal, R., 2000. Physical management of soils of the tropics: Priorities for the 21st century. Soil Science. 165(3), 191-207.

Lamorski, K., Pachepsky, Y., Sławiński, C., Walczak, R.T., 2008. Using support vector machines to develop pedotransfer functions for water retention of soils in Poland. Soil Science Society of American Journal. 72(5), 1243-1247.

Le, V.K., 2003. Physical fertility of typical Mekong delta soils (Vietnam) and land suitability assessment for alternative crops with rice cultivation. Ph.D Dissertation Thesis, Ghent University, Ghent University, 464 pp.

Leenhardt, D., 1995. Errors in the estimation of soil water properties and their propagation through a hydrological model. Soil Use and Management. 11(1), 15-21.

Leij, F.J., Romano, N., Palladino, M., Schaap, M.G., Coppola, A., 2004. Topographical attributes to predict soil hydraulic properties along a hillslope transect. Water Resources Research. 40(2), n/a-n/a.

Li, Y., Šimůnek, J., Jing, L., Zhang, Z., Ni, L., 2014. Evaluation of water movement and water losses in a direct-seeded-rice field experiment using Hydrus-1D. Agricultural Water Management. 142(0), 38-46.

Li, Y., Šimůnek, J., Zhang, Z., Jing, L., Ni, L., 2015. Evaluation of nitrogen balance in a direct-seeded-rice field experiment using Hydrus-1D. Agricultural Water Management. 148(0), 213-222.

Lilly, A., 2000. The relationship between field-saturated hydraulic conductivity and soil structure: development of class pedotransfer functions. Soil Use and Management. 16(1), 56-60.

Lilly, A., Lin, H., 2004. Using soil morphological attributes and soil structure in pedotransfer functions. In: Pachepsky, Y., Rawls, W.J. (Eds.), Developments in Soil Science. Elsevier, pp. 115-141.

Lilly, A., Nemes, A., Rawls, W.J., Pachepsky, Y.A., 2008. Probabilistic Approach to the Identification of Input Variables to Estimate Hydraulic Conductivity. Soil Science Society of American Journal. 72(1), 16-24.

Lin, H., 2012. Hydropedology: Synergistic Integration of Soil Science and Hydrology. Elsevier, Oxford, UK.

Lin, H. et al., 2006. Hydropedology: Synergistic integration of pedology and hydrology. Water Resources Research. 42(5), W05301.

Lin, H.S., McInnes, K.J., Wilding, L.P., Hallmark, C.T., 1999a. Effects of Soil Morphology on Hydraulic Properties I. Quantification of Soil Morphology. Soil Science Society of American Journal. 63(4), 948-954.

Lin, H.S., McInnes, K.J., Wilding, L.P., Hallmark, C.T., 1999b. Effects of soil morphology on hydraulic properties: II. Hydraulic pedotransfer functions. Soil Science Society of American Journal. 63(4), 955-961.

Linh, T.B., Le Van, K., Van Elsacker, S., Cornelis, W.M., 2014. Effect of cropping system on physical properties of clay soil under intensive rice cultivation. Land Degradation & Development, n/a-n/a.

Linh, T.B. et al., 2015a. Inclusion of upland crops in rice-based rotations affects chemical properties of clay soil. Soil Use and Management. 31(2), 313-320.

Linh, T.B., Sleutel, S., Vo Thi, G., Le Van, K., Cornelis, W.M., 2015b. Deeper tillage and root growth in annual rice-upland cropping systems result in improved rice yield and economic profit relative to rice monoculture. Soil and Tillage Research. 154, 44-52.

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.

Mainuddin, M., Kirby, M., Chu, T.H., 2013. Impact of climate change on rainfed rice and options for adaptation in the lower Mekong Basin. Natural Hazards. 66(2), 905-938.

Mandelbrot, B.B., 1983. The fractal geometry of nature. W.H. Freeman, San Francisco, 460 pp.

Manrique, L.A., Jones, C.A., 1991. Bulk Density of Soils in Relation to Soil Physical and Chemical Properties. Soil Science Society of American Journal. 55(2).

Manrique, L.A., Jones, C.A., Dyke, P.T., 1991. Predicting soil water retention characteristics from soil physical and chemical properties. Communications in Soil Science and Plant Analysis. 22(17-18), 1847-1860.

Mayr, T., Jarvis, N.J., 1999. Pedotransfer functions to estimate soil water retention parameters for a modified Brooks–Corey type model. Geoderma. 91(1–2), 1-9.

McBratney, A.B., Minasny, B., Cattle, S.R., Vervoort, R.W., 2002. From pedotransfer functions to soil inference systems. Geoderma. 109(1–2), 41-73.

McBratney, A.B., Minasny, B., Tranter, G., 2011. Necessary meta-data for pedotransfer functions. Geoderma. 160(3–4), 627-629.

McKeague, J.A., Wang, C., Topp, G.C., 1982. Estimating Saturated Hydraulic Conductivity from Soil Morphology. Soil Science Society of American Journal. 46(6), 1239-1244.

McKenzie, N., Jacquier, D., 1997. Improving the field estimation of saturated hydraulic conductivity in soil survey. Soil Research. 35(4), 803-827.

McKenzie, N., MacLeod, D., 1989. Relationships between soil morphology and soil properties relevant to irrigated and dryland agriculture. Soil Research. 27(2), 235-258.

Mckenzie, N., Smettem, K., Ringrose-Voase, A., 1991. Evaluation of methods for inferring air and water properties of soils from field morphology. Soil Research. 29(5), 587-602.

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. In: Yuren, J. (Ed.), 12th ISCO Conference. Tsinghua University Press: Beijing, Beijing, China, pp. 657-662.

Medeiros, J.C., Cooper, M., Dalla Rosa, J., Grimaldi, M., Coquet, Y., 2014. Assessment of pedotransfer functions for estimating soil water retention curves for the amazon region. Revista Brasileira de Ciência do Solo. 38, 730-743.

Medina, H., Tarawally, M., Valle, A.d., Ruiz, M.a.E., 2002. Estimating soil water retention curve in rhodic ferralsols from basic soil data. Geoderma. 108(3–4), 277-285.

Mekong Delta Plan, 2013. Long-term vision and strategy for a safe, prosperous and sustainable delta.

Mensvoort, M.E.F., Van den Berg, M., 1986. Soil morphology: background informations with special reference to the soils in the Mekong delta, Vietnam, NUFFIC project VH-10, Wageningen Agriculture University, The Netherlands.

Merdun, H., Çınar, Ö., Meral, R., Apan, M., 2006. Comparison of artificial neural network and regression pedotransfer functions for prediction of soil water retention and saturated hydraulic conductivity. Soil & Tillage Research. 90(1–2), 108-116.

Millán, H., González-Posada, M., 2005. Modelling soil water retention scaling. Comparison of a classical fractal model with a piecewise approach. Geoderma. 125(1–2), 25-38.

Minasny, B., Hartemink, A.E., 2011. Predicting soil properties in the tropics. Earth-Science Reviews. 106(1–2), 52-62.

Minasny, B., Mc Bratney, A.B., McKenzie, N.J., Grundy, M.J., 2008a. Predicting soil properties using pedotransfer functions and environmental correlation. In:

McKenzie, N.J., Grundy, M.J., Webster, R., Ringrose-Voase, A.J. (Eds.), Guidelines for surveying soil and land resources, Second Edition. Australian Soil and Land Survey Handbooks Series. CSIRO Publishing, pp. 349-367.

Minasny, B., McBratney, A.B., 2002. The efficiency of various approaches to obtaining estimates of soil hydraulic properties. Geoderma. 107(1–2), 55-70.

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(3–4), 225-253.

Minasny, B., McBratney, A.B., Tranter, G., Murphy, B.W., 2008b. Using soil knowledge for the evaluation of mid-infrared diffuse reflectance spectroscopy for predicting soil physical and mechanical properties. European Journal of Soil Science. 59(5), 960-971.

Mishra, S., Parker, J.C., Singhal, N., 1989. Estimation of soil hydraulic properties and their uncertainty from particle size distribution data. Journal of Hydrology. 108, 1-18.

Moeys, J. et al., 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(7), 2069-2083.

Mondal, M.S. et al., 2015. Simulating yield response of rice to salinity stress with the AquaCrop model. Environmental Science: Processes & Impacts. 17(6), 1118-1126.

Mualem, Y., 1976. A new model for predicting the hydraulic conductivity of unsaturated porous media. Water Resources Research. 12(3), 513-522.

Mucherino, A., Papajorgji, P.J., Pardalos, P.M., 2009. Data Mining in Agriculture. In: Panos M. Pardalos, D.-Z.D. (Ed.). Springer Optimization and Its Applications. Springer New York, pp. 272.

Nebel, Á.L.C. et al., 2010. Pedotransfer functions related to spatial variability of water retention attributes for lowland soils. Revista Brasileira de Ciência do Solo. 34, 669-680.

Nemes, A., Czinege, E., Farkas, C., 2010. Use of simulation modeling and pedotransfer functions to evaluate different irrigation scheduling scenarios in a heterogeneous field 19th World Congress of Soil Science, Soil Solutions for a Changing World Brisbane, Australia, pp. 186-189.

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 American Journal. 70(2), 327-336.

Nemes, A., Rawls, W.J., Pachepsky, Y.A., van Genuchten, M.T., 2006b. Sensitivity analysis of the nonparametric nearest neighbor technique to estimate soil water retention. Vadose Zone Journal. 5(4), 1222-1235.

Nemes, A., Roberts, R.T., Rawls, W.J., Pachepsky, Y.A., van Genuchten, M.T., 2008. Software to estimate −33 and −1500 kPa soil water retention using the non-parametric k-Nearest Neighbor technique. Environmental Modelling & Software. 23(2), 254-255.

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 American Journal. 67(4), 1093-1102.

Nemes, A., Wösten, J.H.M., Lilly, A., Oude Voshaar, J.H., 1999. Evaluation of different procedures to interpolate particle-size distributions to achieve compatibility within soil databases. Geoderma. 90(3–4), 187-202.

Nguyen, H.N., Vu, K.T., Nguyen, X.N., 2007. Flooding in Mekong River Delta, Vietnam. 53, Human Development Report Office, UNDP.

Nguyen, V.Q., 2011. Application of the crop water model AquaCrop in the Mekong Delta, Vietnam. Master dissertation Thesis, Ghent University, Gent, Belgium, 64 pp.

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. 27(19), 2721-2735.

Or, D., Wraith, J.M., 2002. Soil water content and water potential relationships. In: Warrick, A.W. (Ed.), Soil physics companion. CRC Press, Boca Raton, FL, pp. 49-82.

Osman, K.T., 2013. Soil degradation, conservation and remediation. Springer Dordrecht Heidelberg, London, 248 pp.

Pachepsky, Y., Rajkai, K., Tóth, B., 2015. Pedotransfer in soil physics: trends and outlook — A review. Agrokémia és Talajtan. 64(2), 339-360.

Pachepsky, Y., Rawls, W., Gimenéz, 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(1–2), 117-126.

Pachepsky, Y., Rawls, W.J., 2004. Development of pedotransfer functions in soil hydrology. Developments in Soil Science, 30. Elsevier, 512 pp.

Pachepsky, Y., Schaap, M.G., 2004. Data mining and exploration techniques. In: Pachepsky, Y., Rawls, W.J. (Eds.), Developments in Soil Science. Elsevier, pp. 21-32.

Pachepsky, Y.A., Rawls, W.J., 1999. Accuracy and reliability of pedotransfer functions as affected by grouping soils. Soil Science Society of American Journal. 63(6), 1748-1757.

Pachepsky, Y.A., Rawls, W.J., 2003. Soil structure and pedotransfer functions. European Journal of Soil Science. 54(3), 443-452.

Pachepsky, Y.A., Rawls, W.J., Lin, H.S., 2006. Hydropedology and pedotransfer functions. Geoderma. 131(3–4), 308-316.

Pachepsky, Y.A., Rawls, W.J., Timlin, D.J., 2013. The current status of pedotransfer functions: their accuracy, reliability, and utility in field- and regional-scale modeling. In: Corwin, D.L., Loague, K., Ellsworth, T.R. (Eds.), Assessment of non-point source pollution in the vadose zone. American Geophysical Union, pp. 223-234.

Pachepsky, Y.A., Timlin, D., Varallyay, G., 1996a. Artificial Neural Networks to Estimate Soil Water Retention from Easily Measurable Data. Soil Science Society of American Journal. 60(3), 727-733.

Pachepsky, Y.A., Timlin, D., Varallyay, G., 1996b. Artificial neural networks to estimate soil water retention from easily measurable data. Soil Science Society of American Journal. 60(3), 727-733.

Pachepsky, Y.A., Timlin, D.J., Rawls, W.J., 2001. Soil Water Retention as Related to Topographic Variables. Soil Science Society of American Journal. 65(6).

Page, A.L., 1982. Methods of soil analysis. Part 2. Chemical and microbiological properties. Soil Science Society of America Book Series. American Society of Agronomy, Soil Science Society of America, Madison, Wisconsin, USA, 1159 pp.

Patil, N., 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. 135(3), 286-294.

Patil, N. et al., 2013. Soil water retention characteristics of black soils of India and pedotransfer functions using different approaches. Journal of Irrigation and Drainage Engineering. 139(4), 313-324.

Perkins, K., Nimmo, J., 2009. High-quality unsaturated zone hydraulic property data for hydrologic applications. Water Resources Research. 45(7), W07417

Phogat, V., Yadav, A.K., Malik, R.S., Kumar, S., Cox, J., 2010. Simulation of salt and water movement and estimation of water productivity of rice crop irrigated with saline water. Paddy Water Environment. 8(4), 333-346.

Pidgeon, J.D., 1972. Measurement and prediction of available water capacity of ferrallitic soils in Uganda. Soil Science. 23, 431-441.

Pulido Moncada, M., 2014. Integrated assessment of soil structural quality. PhD Thesis, Ghent Univesity, Gent, Belgium, 209 pp.

Pulido Moncada, M., Gabriels, D., Cornelis, W.M., 2014. Data-driven analysis of soil quality indicators using limited data. Geoderma. 235–236, 271-278.

R Core Team, 2014. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Viena, Austria.

Rajkai, K., Várallyay, G., 1992. Estimating soil water retention from simpler properties by regression techniques. In: van Genuchten, M.T., Leij, F.J., Lund, L.J. (Eds.), Proceedings of International Workshop on Indirect Methods for Estimating the Hydraulic Properties of Unsaturated Soils. University of California: Riverside, CA, Riverside, CA, pp. 417-426.

Ranatunga, K., Nation, E.R., Barratt, D.G., 2008. Review of soil water models and their applications in Australia. Environmental Modelling & Software. 23(9), 1182-1206.

Rawls, W.J., Brakensiek, D.L., 1982. Estimating soil water retention from soil properties. Journal American Society of Civil Engineers, Irrigation and Drainage Division. 108(2), 166-171.

Rawls, W.J., Brakensiek, D.L., 1985. Prediction of Soil Water Properties for Hydrologic Modeling. In: Jones, E.B., Wards, T.J. (Eds.), Watershed management in the Eighties. American Society of Civil Engineers, Denver, Colorado, United States, pp. 293-299.

Rawls, W.J., Brakensiek, D.L., Saxton, K.E., 1982. Estimation of soil-water properties. Transactions of the American Society of Agricultural Engineers. 25, 1316-1328.

Rawls, W.J., Gimenez, D., Grossman, R., 1998. Use of soil texture, bulk density and slope of the water retention curve to predict saturated hydraulic conductivity. Transactions of the American Society of Agricultural Engineers. 41(4), 983-988.

Rawls, W.J., Gish, T.J., Brakensiek, D.L., 1991. Estimating soil water retention from soil physical properties and characteristics. In: Stewart, B.A. (Ed.), Advances in Soil Science. Advances in Soil Science. Springer New York, pp. 213-234.

Rawls, W.J., Nemes, A., Pachepsky, Y., 2004. Effect of soil organic carbon on soil hydraulic properties. In: Pachepsky, Y., Rawls, W.J. (Eds.), Developments in Soil Science. Elsevier, pp. 95-114.

Rawls, W.J., Pachepsky, Y.A., 2002. Soil consistence and structure as predictors of water retention. Soil Science Society of American Journal. 66(4), 1115-1126.

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(1–2), 61-76.

Reichert, J.M. et al., 2009. Estimation of water retention and availability in soils of Rio Grande do Sul. Revista Brasileira de Ciência do Solo. 33, 1547-1560.

Reynolds, W.D. et al., 2007. Land management effects on the near-surface physical quality of a clay loam soil. Soil & Tillage Research. 96(1–2), 316-330.

Rezaei, M. et al., 2016. Sensitivity of water stress in a two-layered sandy grassland soil to variations in groundwater depth and soil hydraulic parameters. Hydrology and Earth System Sciences. 20(1), 487-503.

Richards, L.A., 1931. Capillary conduction of liquids through porous mediums. Journal of Applied Physics. 1(5), 318-333.

Rijsberman, F.R., 2006. Water scarcity: Fact or fiction?. Agricultural Water Management. 80(1–3), 5-22.

Rohdenburg, H., Diekkrüger, B., Bork, H.R., 1986. Deterministic hydrological site and catchment models for the analysis of agroecosystems. Catena. 13(1–2), 119-137.

Romano, N., Chirico, G.B., 2004. The role of terrain analysis in using and developing pedotransfer functions. In: Pachepsky, Y., Rawls, W.J. (Eds.), Developments in Soil Science. Elsevier, pp. 273-294.

Romano, N., Palladino, M., 2002. Prediction of soil water retention using soil physical data and terrain attributes. Journal of Hydrology. 265(1–4), 56-75.

Salchow, E., Lal, R., Fausey, N.R., Ward, A., 1996. Pedotransfer functions for variable alluvial soils in southern Ohio. Geoderma. 73(3–4), 165-181.

Santos, W.J.R., Curi, N., Silva, S.H.G., Araújo, E.F., 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. 37(1), 49-60.

Santra, P., Das, B.S., 2008. Pedotransfer functions for soil hydraulic properties developed from a hilly watershed of Eastern India. Geoderma. 146(3–4), 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 American Journal. 70(5), 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 American Journal. 50(4), 1031-1036.

Saxton, K.E., Willey, P.H., 2006. The SPAW model for agricultural field and pond hydraulic simulation. In: Singh, V.P., Frevert, D.K. (Eds.), Watershed models. CRC Press, Boca Raton, pp. 401-435.

Scanlon, B.R., Keese, K., Reedy, R.C., Simunek, J., Andraski, B.J., 2003. Variations in flow and transport in thick desert vadose zones in response to paleoclimatic forcing (0–90 kyr): Field measurements, modeling, and uncertainties. Water Resources Research. 39(7).

Schaap, M.G., 2004. Accuracy and uncertainty in PTF predictions. In: Pachepsky, Y., Rawls, W.J. (Eds.), Developments in Soil Science. Elsevier, 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, New York, pp. 1145-1150.

Schaap, M.G., Leij, F.J., 1998. Using neural networks to predict soil water retention and soil hydraulic conductivity. Soil & Tillage Research. 47(1–2), 37-42.

Schaap, M.G., Leij, F.J., van Genuchten, M.T., 1998. Neural network analysis for hierarchical prediction of soil hydraulic properties. Soil Science Society of American Journal. 62(4), 847-855.

Schaap, M.G., Leij, F.J., van Genuchten, M.T., 1999. A bootstrap-neural network approach to predict soil hydraulic parameters. In: Van Genuchten, M.T., Leij, F.J., Wu, L. (Eds.), International Workshop: Characterization and Measurements of the hydraulic properties of saturated porous media, University of California, Riverside, USA, pp. 1237-1250.

Schaap, M.G., Leij, F.J., van Genuchten, M.T., 2001. Rosetta: a computer program for estimating soil hydraulic parameters with hierarchical pedotransfer functions. Journal of Hydrology. 251(3–4), 163-176.

Scheinost, A.C., Sinowski, W., Auerswald, K., 1997. Regionalization of soil water retention curves in a highly variable soilscape, I. Developing a new pedotransfer function. Geoderma. 78(3–4), 129-143.

Sharma, S.K., Mohanty, B.P., Zhu, J., 2006. Including topography and vegetation attributes for developing pedotransfer functions. Soil Science Society of American Journal. 70(5), 1430-1440.

Shmilovici, A., 2010. Support Vector Machines. In: Maimon, O., Rokach, L. (Eds.), Data Mining and Knowledge Discovery Handbook, 2nd edition. Springer-Verlag New York, Inc., pp. 1285.

Shrestha, S., Deb, P., Bui, T.T.T., 2016. Adaptation strategies for rice cultivation under climate change in Central Vietnam. Mitigation and Adaptation Strategies for Global Change. 21(1), 15-37.

Shrestha, S., Thin, N.M.M., Deb, P., 2014. Assessment of climate change impacts on irrigation water requirement and rice yield for Ngamoeyeik Irrigation Project in Myanmar. Journal of Water and Climate Change. 5(3), 427-442.

Shwetha, P., Varija, K., 2013. Soil water-retention prediction from pedotransfer functions for some Indian soils. Archives of Agronomy and Soil Science. 59(11), 1529-1543.

Šimůnek, J., Genuchten, M.T.v., Šejna, M., 2008. Development and Applications of the HYDRUS and STANMOD Software Packages and Related Codes. Vadose Zone Journal. 7(2).

Šimůnek, J., van Genuchten, M.T., Šejna, M., 2012. HYDRUS: Model use, calibration, and validation. Transactions of the ASABE 55(4), 1261-1274.

Skalová, J., Čistý, M., Bezák, J., 2011. Comparison of three regression models for determining water retention curves. Journal of Hydrology and Hydromechanics. 59(4), 275-284.

Smola, A., Schölkopf, B., 2004. A tutorial on support vector regression. Statistics and Computing. 14(3), 199-222.

Soil Survey Staff, 1975. Soil Taxonomy, Soil Conservation Service, USDA, Washington, DC., USA.

Soil Survey Staff, 1990. Keys to Soil Taxonomy, 4th edition, Virginia Polytechnic Institute and State University, Soil Conservation Service, USDA

Washington D.C, USA.

Soil Survey Staff, 1997. National characterization data, Soil Survey Laboratory, National Soil Survey Center, and Natural Resources Conservation Service, Lincoln, NE.

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(3), 426-437.

Steduto, P. et al., 2008. AquaCrop: a new model for crop prediction under water deficit conditions, International Conference 'Drought management: Scientific and technological innovation'. Options Mediterraneennes, Series A, Zaragova, Spain, pp. 285-292.

Suprayogo, D., Widianto, Cadisch, G., Noordwijk, M.V., 2003. A Pedotransfer Resource Database (PTFRDB) for tropical soils: test with the water balance of WaNuLCAS. In: Post, D.A. (Ed.), MODSIM 2003. Modelling and Simulation Society of Australia and New Zealand Inc. 2003: Canberra, Australia, Townsville, Australia, pp. 584-589.

Sys, C., Van Ranst, E., Debaveye, J., 1991. Land evaluation. Part 1. Brussels, General Administration for Development Cooperation.

Tan, X., Shao, D., Gu, W., Liu, H., 2015. Field analysis of water and nitrogen fate in lowland paddy fields under different water managements using HYDRUS-1D. Agricultural Water Management. 150(0), 67-80.

Tan, X., Shao, D., Liu, H., 2014. Simulating soil water regime in lowland paddy fields under different water managements using HYDRUS-1D. Agricultural Water Management. 132(0), 69-78.

Tessier, D., Lajudie, A., Petit, J.C., 1992. Relation between the macroscopic behavior of clays and their microstructural properties. Applied Geochemistry. 7, Supplement 1(0), 151-161.

Tietje, O., Tapkenhinrichs, M., 1993. Evaluation of pedo-transfer functions. Soil Science Society of America Journal. 57(4), 1088-1095.

Tomasella, J., Hodnett, M., 2004. Pedotransfer functions for tropical soils. In: Pachepsky, Y., Rawls, W.J. (Eds.), Developments of Pedotransfer Functions in Soil Hydrology. Elsevier, pp. 415-429.

Tomasella, J., Hodnett, M.G., 1998. Estimating soil water retention characteristics from limited data in Brazilian Amazonia. Soil Science. 163(3), 190-202.

Tomasella, J., Hodnett, M.G., Rossato, L., 2000. Pedotransfer functions for the estimation of soil water retention in brazilian soils. Soil Science Society of American Journal. 64(1), 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 American Journal. 67(4), 1085-1092.

Toumi, J. et al., 2016. Performance assessment of AquaCrop model for estimating evapotranspiration, soil water content and grain yield of winter wheat in Tensift Al Haouz (Morocco): Application to irrigation management. Agricultural Water Management. 163, 219-235.

Tran, K.T., 1998. Pedology, Lecture Notes (in Vietnamese). College of Agriculture. Can Tho University.

Tran, T.B., Bach, T.S., Miller, F., 2007. Challenges to sustainable development in the Mekong delta: Regional and national policy issues and research need. The Sustainable Mekong Research Network (Sumernet), Bangkok, Thailand, 202 pp.

Tranter, G., McBratney, A.B., Minasny, B., 2009. Using distance metrics to determine the appropriate domain of pedotransfer function predictions. Geoderma. 149(3–4), 421-425.

Tranter, G. et al., 2007. Building and testing conceptual and empirical models for predicting soil bulk density. Soil Use and Management. 23(4), 437-443.

Tranter, G., Minasny, B., McBratney, A.B., Rossel, R.A.V., Murphy, B.W., 2008. Comparing Spectral Soil Inference Systems and Mid-Infrared Spectroscopic Predictions of Soil Moisture Retention. Soil Science Society of American Journal. 72(5), 1394-1400.

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 American Journal. 73(5), 1443-1452.

Tyler, S.W., Wheatcraft, S.W., 1990. Fractal processes in soil water retention. Water Resources Research. 26(5), 1047-1054.

UN-Water, 2013a. UN-Water Factsheet on Water and Climate change. UN-Water Organization of the United Nations, Geneva, Switzerland.

UN-Water, 2013b. UN-Water Factsheet Water for Food UN-Water Organization of the United Nations, Geneva, Switzerland.

UN, 2010. World Population Prospects: The 2010 Revision., Population Division. Department of Economic and Social Affairs., New York.

Valipour, M., Banihabib, M.E., Behbahani, S.M.R., 2012. Monthly Inflow Forecasting using Autoregressive Artificial Neural Network. Journal of Applied Sciences. 12(20), 2139-2147.

Valipour, M., Banihabib, M.E., Behbahani, S.M.R., 2013. Comparison of the ARMA, ARIMA, and the autoregressive artificial neural network models in forecasting the monthly inflow of Dez dam reservoir. Journal of Hydrology. 476(0), 433-441.

van Dam, J.C., Groenendijk, P., Hendriks, R.F.A., Kroes, J.G., 2008. Advances of Modeling Water Flow in Variably Saturated Soils with SWAP. Vadose Zone Journal. 7(2).

van den Berg, M., Klamt, E., van Reeuwijk, L.P., Sombroek, W.G., 1997. Pedotransfer functions for the estimation of moisture retention characteristics of Ferralsols and related soils. Geoderma. 78(3–4), 161-180.

van Genuchten, M.T., 1980. A Closed-form Equation for Predicting the Hydraulic Conductivity of Unsaturated Soils. Soil Science Society of American Journal. 44(5), 892-898.

Vapnik, V.N., 1995. The nature of statistical learning theory. Springer-Verlag New York, Inc., 188 pp.

Vaz, C.M.P. et al., 2005. Validation of the Arya and Paris Water Retention Model for Brazilian Soils. Soil Science Society of American Journal. 69(3).

Vepraskas, M.J., Heitman, J.L., Austin, R.E., 2009. Future directions for hydropedology: quantifying impacts of global change on land use. Hydrology and Earth System Sciences. 13(8), 1427-1438.

Vereecken, H., Diels, J., Van Orshoven, J., Feyen, J., Bouma, J., 1992. Functional Evaluation of Pedotransfer Functions for the Estimation of Soil Hydraulic Properties. Soil Science Society of American Journal. 56(5), 1371-1378.

Vereecken, H., Herbst, M., 2004. Statistical regression. In: Pachepsky, Y., Rawls, W.J. (Eds.), Developments in Soil Science. Elsevier, pp. 3-19.

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(6), 389-403.

Vereecken, H. et al., 2010. Using pedotransfer functions to estimate the van Genuchten–Mualem soil hydraulic properties: A review. Vadose Zone Journal. 9(4), 795-820.

Vo, Q.M., 2006. Setting up the fertility capability classification system for intensive rice areas in the Mekong delta (in Vietnamese). PhD thesis Thesis, Can Tho University, Can Tho University, Vietnam.

Vo, T.G., Nguyen, M.H., 2012. Aquaculture and Agricultural production in the Mekong Delta and its effects on nutrient pollution of soil and water. In: Renaud, F.G., Kuenzer, C. (Eds.), The Mekong Delta System - Interdisciplinary analyses of a River Delta. Springer Environmental Science and Engineering, Springer Dordrecht Heidelberg New York London.

Walkley, A., Black, I.A., 1934. An examination of the degtjareff method for determining soil organic matter, and a proposed modification of the chromic acid titration method. Soil Science. 37(1), 29-38.

Williams, J., Prebble, R., Williams, W., Hignett, C., 1983. The influence of texture, structure and clay mineralogy on the soil moisture characteristic. Australian Journal of Soil Research. 21(1), 15-32.

Williams, J., Ross, P., Bristow, K., 1992. Prediction of the Campbell water retention function from texture, structure, and organic matter. In: Van Genuchten, M.T., Leij, F.J., Lund, L.J. (Eds.), Indirect methods for Estimating the Hydraulic Properties of Unsaturated Soils. University of California, Riverside, CA, University of California, Riverside, pp. 427-442.

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(3–4), 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(3–4), 169-185.

Wösten, J.H.M., Nemes, A., 2004. Pedotransfer functions for Europe. In: Pachepsky, Y., Rawls, W.J. (Eds.), Developments in Soil Science. Elsevier, pp. 431-435.

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(3–4), 123-150.

Wösten, J.H.M., Schuren, C.H.J.E., Bouma, J., Stein, A., 1990. Use of practical aspects of soil behaviour to evaluate different methods to generate soil hydraulic functions. Hydrological Processes. 4(4), 299-310.

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. 195–196(0), 79-86.

Yurtseven, E., Šimůnek, J., Avci, S., Öztürk, H.S., 2013. Comparison of HYDRUS-1D Simulations and Ion (Salt) Movement in the Soil Profile Subject to Leaching. In: Šimůnek, J., van Genuchten, M.T., Kodešová, R. (Eds.), the 4th International Conference "HYDRUS Software Applications to Subsurface Flow and Contaminant Transport Problems". Dept. of Soil Science and Geology, Czech University of Life Sciences, Prague, Czech Republic, pp. 395-404.

Zacharias, S., Wessolek, G., 2007. Excluding organic matter content from pedotransfer predictors of soil water retention. Soil Science Society of American Journal. 71(1), 43-50.

APPENDICES

Appendix A. Soil and geological information (e.g., land used types, soil types, horizons, depth of horizons, geological units) of soil samples collected in the framework of the thesis (training data set with N = 160 samples). The meaning of numbers of geological units can be referred in Fig. 2-2.

Sample	Sampling location	Position (UTM) x	Position (UTM) y	current land use/vegetation	horizons	depth (cm)	FAO legend in the soil map	Soil type (FAO)	Geological unit
1	Phung Hiep,			paddy rice	Ap	0-20		Orthi Thionic	
2	Hau Giang	565809	1079433	(after harvest)	Bg	20-50	FLt(oen-j)u	Fluvisols	2
3	Phung Hiep,			upland crop,	Ap	0-15		Orthi Thionic	
4	Hau Giang	567760	1080024	leafy vegetable	Bg	15-30	GLt(oen-j)u	Gleysols	2
5	Phung Hiep,			paddy rice	Ap	0-15		Proto Thionic	
6	Hau Giang	565775	1079289	(rice on the field)	B	15-40	FLt(oen-i)u	Fluvisols	2
7	Chau Thanh,			paddy rice	Ap	0-20			
8	Hau Giang	581625	1101122	(rice on the field)	B	20-40	GLmf	Gleysols	5
9	Chau Thanh,			upland crop/	Ap	0-10			
10	Hau Giang	581585	1101144	maize	B	10-20	GLmf	Gleysols	5
11	Giong Rieng,	549348	1089595	paddy rice	Ap	0-20		Proto Thionic	
12	Kien Giang			(rice on field)	Bg	20-40	GLt(pen)d	Gleysols	2
13	Giong Rieng,			paddy rice	Ap	0-20		Orthi Thionic	
14	Kien Giang	525760	1087840	(rice on the field)	B	20-50	FLt(pep)us	Fluvisols	2
15	Giong Rieng,			upland crop,	Ap	0-25		Orthi Thionis	
16	Kien Giang	525744	1087845	bitter melon	B	25-55	FLt(pep)us	Fluvisols	2
17	Vi Thanh, Hau	543182	1076133	upland crop,	Ap	0-50	FLt(oen-j)u	Orthi Thionic	2

18	Giang			pineapple	B	50-70		Fluvisols	
19	Cho Moi,			upland crop,	Ap	1-30			
20	An Giang	558777	154414	maize	B	30-60	GLmf	Gleysols	5
21	Cho Moi,			paddy rice	Ap	0-25			
22	An Giang	558502	1154171	(after harvest)	B	25-40	GLmf	Gleysols	5
23	Cho Moi,			upland crop,	Ap	0-20			
24	An Giang	560545	1154327	maize	Bg	20-40	Fleg	Fluvisols	5
25	Cho Moi,			paddy rice	Ap	0-15			
26	An Giang	560848	1154510	(after harvest)	Bg	15-30	Fleg	Fluvisols	5
27	Cho Moi,			paddy rice	Ap	0-20		Proto Thionic	
28	An Giang	556727	1154074	(after sowing)	B	20-50	GLt(pen)u	Gleysols	5
29	Cho Moi,			upland crop,	Ap	0-15		Proto Thionic	
30	An Giang	556727	1154074	maize	Bg	15-45	GLt(pen)u	Gleysols	5
31	Tinh Bien,			paddy rice	Ap	0-15			
32	An Giang	502607	1164488	(rice on the field)	B	15-30	ARho	Arenosols	4
33	Tinh Bien,			paddy rice	Ap	0-15			
34	An Giang	502481	1158108	(rice on the field)	B	15-30	PDdo	Acrisols	4
35	Tinh Bien,			paddy rice	A	0-15			
36	An Giang	501878	1156909	(rice on the field)	Bg	15-35	PDdo	Acrisols	4
37	Tri Ton,			paddy rice	Ap	0-10			
38	An Giang	501195	1145384	(rice on the field)	B	10-20	LPdo	Acrisols	4
39	Tri Ton,				A	0-10		Thionic	
40	An Giang	507386	1142189	natural grass	Bt	10-40	GLt(oen)u	Gleysols	2
41	Long Xuyen,			paddy rice	Ap	0-10			
42	An Giang	548572	1149989	(after harvest)	Bt	10-25	LUvd	Luvisols	5
43	Long Xuyen,			upland crop,	Ap	0-15			
44	An Giang	548538	1150020	maize	Bt	15-40	LUvd	Luvisols	5
45	Lai Vung,	563441	1135155	upland field,	Ap	0-20	LUgr	Luvisols	5

46	Dong Thap			bitter melon	Bt	20-40			
47	Lai Vung,			paddy rice	Ap	0-15			
48	Dong Thap	564433	1135012	(rice on the field)	Bt	15-30	LUgr	Luvisols	5
49	Duyen Hai,				A	0-7			
50	Tra Vinh	654713	1061656	natural grass	Bg	7-25	Argd	Arenosols	6
51	Duyen Hai,			upland crop,	A	0-10			
52	Tra Vinh	654699	1061759	melon	B	10-20	Argd	Arenosols	6
53	Duyen Hai,			upland crop,	Ap	0-20		Orthi Thionic	
54	Tra Vinh	644243	1061090	flower	Bg	20-40	FLt(pep)ds	Fluvisols	6
55	Dinh An,			paddy rice	Ap	0-20			
56	Tra Vinh	642954	1064151	(rice on the field)	Bg	20-40	Flegs	Fluvisols	6
57	Tieu Can,			paddy rice	Ap	0-20			
58	Tra Vinh	634497	1176038	(rice on the field)	B	20-35	ARgd	Arenosols	6
59	Tieu Can,			upland crop,	Ap	0-15			
60	Tra Vinh	634476	1176004	bitter melon	B	15-30	ARgd	Arenosols	6
61	Binh Minh,			paddy rice	Ap	0-15			
62	Vinh Long	596245	1105813	(rice on the field)	Bg	15-30	Flem	Fluvisols	5
63	Binh Minh,			upland crop,	Ap	0-30			
64	Vinh Long	596245	1105813	leafy vegetable	Bg	30-50	Flem	Fluvisols	5
65	Binh Minh,			upland crop,	Ap	0-15			
66	Vinh Long	596281	1105822	grapefruit orchard	Btg	15-50	Flem	Fluvisols	5
67	Binh Minh,			upland crop,	Ap	0-20		Proto Thionic	
68	Vinh Long	589015	1111455	water melon	Bg	20-30	FLt(pen)u	Fluvisols	5
69	Binh Minh,			upland crop,	Ap	0-25		Proto Thionic	
70	Vinh Long	584252	1111746	grapefruit orchard	Bg	25-45	FLt(pen)u	Fluvisols	5
71	Binh Tan,			paddy rice	Ap	0-10			
72	Vinh Long	585860	1114400	(rice on the field)	Bt	10-25	ALgr	Fluvisols	5
73	Binh Tan,	585830	1114495	upland crop,	Ap	0-15	ALgr	Fluvisols	5

	Vinh Long			mungbean +					
74				maize	Bt	15-40			
75	Binh Tan,			upland crop,	Ap	0-25			
76	Vinh Long	585849	1114475	grapefruit orchard	Bt	25-40	ALgr	Fluvisols	5
77	Cu Lao Dung,			upland crop,	Ap	0-20			
78	Soc Trang	621334	1075237	sugarcane	Bg	20-45	FLes	Fluvisols	6
79	Cu Lao Dung,			upland crop,	Ap	0-15			
80	Soc Trang	626877	1065410	maize	Bg	15-30	Flegs	Fluvisols	6
81	Long Phu,			upland crop,	Ap	0-25			
82	Soc Trang	624351	1063215	maize	Bt	25-45	LUvdhs	Luvisols	6
83	Long Phu,			paddy rice	Ap	0-20			
84	Soc Trang	624193	1063126	(after harvest)	Bt	20-35	LUvdhs	Luvisols	6
85	Vinh Chau,			upland crop,	A1	0-20			
86	Soc Trang	631172	1051553	longan garden	A2	20-40	ARgd	Arenosols	6
87	Vinh Chau,			paddy rice	Ap	0-15			
88	Soc Trang	624308	1035095	(rice on the field)	B	15-40	ARgd	Arenosols	6
89				upland crop,	Ap	0-15			
	Thanh Tri,			okras + bitter					
90	Soc Trang	578334	1038551	melon	B	15-35	GLehs	Gleysols	2
91	Thanh Tri,			paddy rice	Ap	0-10			
92	Soc Trang	578334	1038551	(after harvest)	Bg	10-30	GLehs	Gleysols	2
93	Thanh Tri,			upland crop,	Ap	0-15			
94	Soc Trang	580025	1041710	maize	Bg	15-30	GLuhhs	Gleysols	2
95	Thanh Tri,			paddy rice	Ap	0-20			
96	Soc Trang	580210	1041726	(rice on the field)	B	20-30	GLuhhs	Gleysols	2
97	Nga Nam,			paddy rice	Ap	0-10			
98	Soc Trang	572871	1057656	(rice on the field)	Bg	10-25	FLt(oep)uhs	Fluvisols	2
99	Nga Nam,	567559	1052443	paddy rice	Ap	0-20	FLt(oep-	Fluvisols	2

100	Soc Trang			(rice on the field)	Bj	20-45	j)uhs		
101	Moc Hoa,			paddy rice	Aph	0-17			
102	Long An	601842	1196042	(after harvest)	Bg	17-50	PT(ap)en	Plinthosols	3
103	Moc Hoa,			rotation of paddy	Ap	0-20			
104	Long An	601842	1196042	rice + water melon	Bg	20-50	PT(ap)en	Plinthosols	3
105	Moc Hoa,			paddy rice	Ap	0-20			
106	Long An	599741	1197444	(fallow period)	Bg	20-30	PT(ap)ep	Plinthosols	3
107	Moc Hoa,			upland crop,	Ap	0-10			
108	Long An	599661	1197260	sweet potato	Bg	10-30	PT(ap)ep	Plinthosols	3
109	Vinh Hung,			paddy rice	Ap	0-26			
110	Long An	585075	1202574	(rice on the field)	Bg	26--50	PTau	Plinthosols	3
111				upland crop,	Ap	0-10			
	Vinh Hung,		1202574	water morning					
112	Long An	585075		glory	Bg	10-40	PTau	Plinthosols	3
113	Vinh Hung,			upland crop,	Ap	0-15			
114	Long An	585075	1202574	mango orchard	B	15-30	PTau	Plinthosols	3
115	Vinh Hung,			paddy rice	Ap	0-13			
116	Long An	585075	1202574	(rice on the field)	Bg	13-65	PTau	Plinthosols	3
117	Moc Hoa,			paddy rice	Ap	0-10		Proto Thionic	
118	Long An	610951	1179999	(rice on the field)	Bt	10-30	GLt(pen)u	Gleysols	2
119	Ninh Kieu,			upland crop,	Ap	0-15			
120	Can Tho	583940	1108734	mango orchard	Bg	15-30	FLeg	Fluvisols	5
121	Ba Tri,			paddy rice	Ap	0-20		Proto Thionic	
122	Ben Tre	673137	1125765	(rice on the field)	B	20-40	FLt(pen)ds	Fluvisols	5
123	Cang Long,			paddy rice	Ap	0-30			
124	Tra Vinh	637454	1103736	(rice on the field)	Bt	30-55	LUgr	Fluvisols	5
125	Cang Long,	637454	1103736	grass + banana	A	0-10	LUgr	Fluvisols	5

126	Tra Vinh				Bg	10-20			
127	Tam Binh,			paddy rice	Ap	0-18			
128	Vinh Long	597731	1116742	(rice on the field)	Btg	18-30	GLmf	Gleysols	5
129	Tam Binh,			orchard of mango	Ap	0-10			
130	Vinh Long	597731	1116742	+ durian	B	10-20	GLmf	Gleysols	5
131	Tam Binh,			paddy rice	Ap	0-20		Proto Thionic	
132	Vinh Long	602345	1113217	(rice on the field)	Btg	20-40	GLt(oen-j)u	Gleysols	5
133	Tam Binh,			upland crop,	Ap	0-15		Proto Thionic	
134	Vinh Long	602400	1113197	grapefruit orchard	Bg	15-30	GLt(oen-j)u	Gleysols	5
135	Tam Binh,			paddy rice	Ap	0-20			
136	Vinh Long	607533	1110631	(rice on the field)	Bt	20-35	GLmf	Gleysols	5
137	Tam Binh,				Ap	0-25			
138	Vinh Long	607480	1110579	durian orchard	Bt	25-40	GLmf	Gleysols	5
139	Long Ho, Vinh			paddy rice	Ap	0-18		Thionic	
140	Long	612296	1121726	(rice on the field)	B	18-35	FLt(pen)u	Fluvisols	5
141	Long Ho, Vinh			upland crop,	Ap	0-35		Thionic	
142	Long	61225	1121731	mango orchard	B	35-55	FLt(pen)u	Fluvisols	5
143	Tra On,			paddy field	Ap	0-30			
144	Vinh Long	607723	1101015	(rice on the field)	Btg	30-50	LUgr	Luvisols	5
145	Tra On,			upland crop,	Ap	0-20			
146	Vinh Long	607687	1101028	grapefruit orchard	Bg	20-40	GLuh	Gleysols	5
147	Tra On,			paddy rice	Ap	0-30			
148	Vinh Long	613430	1100139	(rice on the field)	Bt	30-40	GLuh	Gleysols	5
149	Tra On,				Ap	0-15			
150	Vinh Long	613421	1100133	citrus orchard	Bb	15-45	GLuh	Gleysols	5
151	Vung Liem,	620789	1105996	paddy rice	Ap	0-10	FLt(pen)u	Thionic	5

152	Vinh Long			(rice on the field)	Bg	10-35		Fluvisols	
153	Vung Liem,				Ap	0-25		Thionic	
154	Vinh Long	620789	1105996	banana + coconut	B	25-45	FLt (pen)u	Fluvisols	5
155	Cang Long,			paddy rice	Ap	0-30			
156	Tra Vinh	632727	1097605	(rice on the field)	Bt	30-60	ALgr	Alluvisols	6
157	Hung Phu,			paddy rice	Ap	0-25		Gleyic	
158	Can Tho	586763	1107338	(rice on the field)	B	25-45	FLeg	Fluvisols	5
159	Hung Phu,			mango orchard	Ap	0-40		Gleyic	
160	Can Tho	586763	1107338	(>20 years old)	Bt	40-60	FLeg	Fluvisols	5

Appendix B. Soil and geological information (e.g., land used type, soil type, horizon, depth of horizon, geological unit) of soil samples from 10 soil profiles taken from the study of Le (2003). The meaning of numbers of geological units can be referred in Fig. 2-2.

Profile location	Sample	Land used type	Horizon	Depth (cm)	GWT (cm)	Soil group	FAO/UNESCO soil type	Geological unit
Chau Thanh, Tra Vinh	1	two rice crops per year	Ap	0-40	140	well-developed alluvial soil	Gleysols	6
	2		Bg1	40-80				
	3		Bg2	80-120				
Vung Liem, Vinh Long	4	three rice crops per year	Ap	0-20	110	well-developed alluvial soil	Gleysols	6
	5		Bg1	20-62				
	6		Bg2	62-105				
Cai Lay, Tien Giang	7	three rice crops per year	Ap	0-15	150	well-developed alluvial soil	Gleysols	5
	8		Bg1	15-65				
	9		Bg2	65-115				
Cao Lanh, Dong Thap	10	three rice crops per year	Ap	0-15	150	well-developed alluvial soil	Gleysols	5
	11		Bg1	15-65				
	12		Bg2	65-115				
Tan An, Long An	13	three rice crops per year	Ap	0-25	120	well-developed alluvial soil	Gleysols	5
	14		Bg1	40-75				
	15		Bg2	75-120				
Phung Hiep, Hau Giang	16	wild land	Ah	0-25	120	severe acid sulphate soil	Orthi-Thionic Gleysol	2
	17		Bgj	45-90				
	18		Cj	90-120				
Vinh Chau, Soc Trang	19	sugarcane, guava, and other vegetable	Ap	0-45	120	saline soil	Salic Fluvisols	6
	20		Ab	45-95				
	21		Cg	95-120				
O Mon, Can Tho	22	three rice crops per year	Ap	0-22	70	cultivated alluvial soil	Gleysols	5
	23		Bg	22-60				

	24		Cr	>60				
Tinh Bien, An Giang	25 26 27	upland rice in rainy season + cash crop in wet season	Ap Bg1 Bg2	0-18 18-83 83-100	did not observed	weathered soils in mountainous area	Acrisols	4
Moc Hoa, Long An	28 29	two rice crops per year	Ap Bg1	0-15 15-55	80	well-developed old alluvial soil	Plinthosols	3

CURRICULUM VITAE

Personal data:

Name: Nguyen Minh Phuong

Birthday: 7th July 1979

Birth place: Bac Lieu province, Vietnam

Nationality: Vietnamese

Office address: Department of Soil Science

College of Agriculture and Applied Biology

Campus 2, Can Tho University

3/2 street, Ninh Kieu district, Can Tho city, Vietnam

Tel: +84 710 3835062

Fax: +84 710 3830814

E-mail: nmphuong@ctu.edu.vn

Education

1997-2001: Engineer in Land Management, Can Tho University, Vietnam.

2004-2006: Master of Science in Physical Land Resources - speciality in Soil Science, Ghent University, Belgium.

2010 up to now: Doctor of Applied Biological Science: Land and Water Management, Ghent University, Belgium

Awards:

- Certificate of Merit - Cantho University, Vietnam (1997 - 2001).

- Price of De Boodt Maselis for outstanding student of Physical Land Resources Programme, Ghent University, Belgium (2006).

Research and professional experience

From 2001 to 2004: Researcher at the Department of Soil Science and Land Management, Faculty of Agriculture and Applied Biology, Can Tho University.

From 2004 to 2006: Post-graduate research experience in the International Course Program Master in Physical Land Resources, option: Soil Science at Ghent University, Belgium

From 2006 to 2010: Researcher and assistant lecturer at the Department of Soil Science, Faculty of Agriculture and Applied Biology, Can Tho University.

From 2010 up to now: Lecturer at the Department of Soil Science, Can Tho University; and doctoral student at the Department of Soil Management, Faculty of Bioscience Engineering, Ghent University, Belgium.

Doctoral training program related to research

Seminar in transferable skills on:

Effective slide design

Workshop "A hands-on Approach to Athena Visual Studio"

A hands-on training: Writing and reviewing scientific literature in applied biological and veterinary science

English presentation skill in Conference

Flames course: Method in Research Design

Post-graduate courses on:

Applied Linear Regression

Flames Summer School in Methodology and Statistics 2013 in KU Leuven, Belgium.

Tutorship

2011: Tutor of the MSc. thesis of Lieselot Hernalsteen, student in Bioscience Engineering. Thesis entitled “Developing pedotransfer functions for a sustainable soil-water management for the Mekong Delta, Vietnam” under the direction of Prof. Dr. ir. Wim M. Cornelis

2011: Tutor of the MSc. thesis of Dereje Tirfessa Woldetensay, student in Physical Land Resources. Thesis entitled "Use of structural soil properties for predicting water retention properties of some Ethiopian soils" under the direction of Prof. Dr. ir. Wim M. Cornelis

Scientific publications

Articles-A1 (peer-reviewed journal in ISI Web of Science)

Phuong Minh Nguyen, Khoa Van Le, Wim Cornelis, 2014. *Using categorical soil structure information to improve soil water retention estimates of tropical delta soils.* Soil Research 52 (5): 443-452.

Phuong Minh Nguyen, Jan de Pue, Khoa Van Le, Wim Cornelis, 2015*. Impact of regression methods on improved effects of soil structure on soil water retention estimates.* Journal of Hydrology 525: 598-606.

Phuong Minh Nguyen, Yves-dady M. Botula, Khoa Van Le, Wim Cornelis, 2015. *Evaluation of soil water retention pedotransfer functions for Vietnamese Mekong Delta soils.* Agricultural Water Management 158: 126-138.

Phuong Minh Nguyen, Amir Haghverdi, Yves-dady Botula, Jan de Pue, Khoa Van Le, Willem Waegeman, Wim Cornelis, 2016. *Comparison of statistical regression and data mining techniques in predicting soil water retention of tropical delta soils.* Biosystems Engineering (In Press).

Articles-P1 (conference proceeding in ISI Web of Science)

Phuong Minh Nguyen, Khoa Van Le, Wim Cornelis, 2014. *Predicting soil water retention characteristics for Vietnam Mekong Delta soils*. Proceedings of FRIEND-Water 2014, Monpellier, October 2014, IAHS Publ. 363.

Conference contributions

Phuong Minh Nguyen, Lieselot Hernalsteen, Khoa Van Le, Yves-dady M. Botula, Wim Cornelis. Poster presentation: *"Using categorical soil structure to improve soil water retention estimates of tropical Mekong Delta soils"*. 4th Eurosoil International Congress, July 2-6, Bari, Italy.

Phuong Minh Nguyen, Khoa Van Le, Yves-dady M. Botula, Wim Cornelis. Poster presentation: "*Improving soil water retention estimates for tropical Mekong*

Delta soils". 18th National Symposium on Applied Biological Sciences. 8th February 2013, Ghent University, Belgium.

Phuong Minh Nguyen, Yves-dady Botula Manyala, Linh Tran Ba, Khoa Van Le, Wim Cornelis, 2014. Poster presentation: *"Evaluation of soil water retention PTFs for tropical Mekong Delta soils"*. 20th World Congress of Soil Science, June 8-13, Jeju, Korea.

Phuong Minh Nguyen, Khoa Van Le, Wim Cornelis, 2014. Oral presentation: "*Predicting soil water retention characteristics for Vietnam Mekong Delta soils"*. FRIEND-Water 2014 Conference, October 7-10, Montpellier, France.

Phuong Minh Nguyen, Yves-dady Botula Manyala, Khoa Van Le, Wim Cornelis, 2016. Oral presentation: *"Evaluation of soil water retention pedotransfer functions for Vietnamese Mekong Delta"*. Day of Belgian Young Soil Scientist, March 16th , Brussels, Belgium.