A Polynomial Chaos Stochastic Model for 2D Breast Microwave Scattering

Ann Franchois
IDLab – INTEC Dept.
Ghent University
Ghent, Belgium
ann.franchois@ugent.be

Karel van der Steen
IDLab – INTEC Dept.
Ghent University
Ghent, Belgium
karel.vandersteen@ugent.be

Simon De Ridder
IDLab – INTEC Dept.
Ghent University
Ghent, Belgium
simon.deridder@ugent.be

Michiel Gossye
IDLab – INTEC Dept.
Ghent University
Ghent, Belgium
michiel.gossye@ugent.be

Martijn Huynen
IDLab – INTEC Dept.
Ghent University
Ghent, Belgium
martijn.huynen@ugent.be

Hendrik Rogier
IDLab – INTEC Dept.
Ghent University
Ghent, Belgium
hendrik.rogier@ugent.be

***Abstract*—A computational model that quantifies the statistical distributions of the electromagnetic field scattered by a breast with random variations in its complex permittivity profile is presented. It involves the construction of a sparse hyperbolically truncated polynomial chaos expansion for the scattered field at each receiving antenna with the modified cross-validation least-angle regression method and requires significantly less deterministic model evaluations than Monte-Carlo sampling. Tumor detectability, via the Cramér-von Mises criterion, is studied for a simplified and a more realistic two-dimensional breast model. In the latter case, the Karhunen-Loéve transform is applied to reduce the stochastic dimensionality resulting from a large number of correlated permittivity values.**

***Index Terms*—microwave tomography, breast cancer, polynomial chaos, sparse regression**

I. Introduction

Microwave tomography is considered a promising alternative to mammography screening, particularly since electromagnetic waves in the microwave range are non-ionizing [1]. In practice, antennas are arranged around the surface of the breast to collect the electric fields resulting from interactions of the breast's complex permittivity profile with different transmitted waves. The data can be processed numerically to provide an image of the breast interior [2]. The present paper focuses on a statistical solution of the forward scattering problem, to study the effects of random variations in the tissues' complex permittivity on the scattered field. Quantifying the statistical distributions can provide new insights on tumor detectability. One way to analyze the statistics is to perform Monte-Carlo (MC) simulations for a large number of permittivity profiles, leading to a huge computational effort when a full-wave forward solver is employed. An alternative is to compute the scattered field from a polynomial chaos (PC) expansion. Whereas in antenna design the number of uncorrelated random variables (RVs) may be small enough to achieve a substantial gain in computational effort [3], this is much less the case in microwave tomography. Therefore we adopt a sparse Hyperbolically Truncated PC expansion (HTPC), constructed by the Modified Cross-Validation (MCV) Least-Angle Regression (LAR) method [4]. We optimize and validate the method by comparing the output distributions of the sparse HTPC expansions and MC simulations performed with a deterministic full-wave solver. Next, we study tumor detectability by comparing the output distributions for profiles with and without a tumor. The underlying idea is to observe the differences between the cumulative distribution functions, accomplished via the Cramér-von Mises criterion [5]. The paper is organized as follows. In Section II the configuration and methods are described. The results are discussed in Section III. Section IV concludes the paper.

II. Methods

Assume a two-dimensional (2D) breast model Ω characterized by a relative complex permittivity $\epsilon_r(\boldsymbol{r})$ that is invariant along the z-direction. It is surrounded by a coupling medium to increase the microwave energy coupled into the breast. As in [6], there are 16 receiving antenna locations $\boldsymbol{r^R}$ on a circle with diameter 0.15m around Ω, which are numbered counterclockwise. Only the first location, $\boldsymbol{r^R}_1 = (0.075\mathrm{m}, 0) = \boldsymbol{r^S} = (\mathrm{x^S}, \mathrm{y^S})$, is used also for transmitting. The source emits a time-harmonic electromagnetic wave at a frequency of 1.1GHz. The multiple receivers measure the total electric field $\boldsymbol{E}$. When Ω is not present, $\boldsymbol{E}$ corresponds to the incident field $\boldsymbol{E^i}$. The presence of Ω influences the total field, which is written as $\boldsymbol{E} = \boldsymbol{E^i} + \boldsymbol{E^s}$, thereby defining the scattered electric field. The source is a line source and generates a Transverse Magnetically (TM) polarized cylindrical wave $\boldsymbol{E^i} = E^i \boldsymbol{u_z}$. Consequently all fields are TM-polarized and are denoted further by their complex amplitudes E, E^i and E^s, where the time-depence $e^{j\omega t}$ is omitted. The exact relation between the permittivity profile $\epsilon_r(\boldsymbol{r})$ and the

scattered field at a receiver location can be expressed by the integral

$$E^s(\boldsymbol{r}^{\boldsymbol{R}}) = k_0^2 \int_\Omega (\epsilon_r(\boldsymbol{r}') - \epsilon_{r,b})\, E(\boldsymbol{r}') G_b(\boldsymbol{r}^{\boldsymbol{R}} - \boldsymbol{r}') d\boldsymbol{r}', \quad (1)$$

with k_0 the propagation constant of vacuum, $\epsilon_{r,b}$ the relative complex permittivity of the coupling medium and G_b the 2D scalar Green's function. Eq. (1) is employed for the exact numerical computation of the scattered field by a Method of Moments (MoM) discretized full-wave forward solver [7]. The speed of the solver is significantly enhanced by the iterative Bi-CGSTAB-FFT technique [8]. The solver is instrumental in building and validating the stochastic model and will be referred to as the deterministic solver.

Consider a rectangular domain $\mathcal{D}$ in the xy-plane with $\Omega \subset \mathcal{D}$. The relative complex permittivity $\epsilon_r(\boldsymbol{r}) = \epsilon_r^{\mathrm{re}}(\boldsymbol{r}) + \mathrm{j}\epsilon_r^{\mathrm{im}}(\boldsymbol{r})$, with re and im the real and imaginary parts, respectively, is discretized on $\mathcal{D}$ into $N = N_x \times N_y$ square cells with centers $\boldsymbol{r}_n, n = 1, ..., N$ and size Δ. Consequently, the permittivity assumes a constant value within each cell, that depends on the type of tissue (or coupling medium) assigned to that cell. The discretized permittivity profile is represented by a real vector $\boldsymbol{\varepsilon} = [\epsilon_1^{\mathrm{re}}, ..., \epsilon_N^{\mathrm{re}}, \epsilon_1^{\mathrm{im}}, ..., \epsilon_N^{\mathrm{im}}]$ of dimension $2N$, where for simplicity the subscript r is omitted. In our examples, we adopt a square domain $\mathcal{D}$ and in view of the Fast Fourier Transform (FFT) implementation, we set the number of cells in each direction equal to $N_x = N_y = 8 \times 2^r + 1$, with r a positive integer. We choose the cell size Δ sufficiently small to secure a discretization error on the scattered fields below 0.3%. The scattered field computed by the deterministic solver then is considered as the exact scattered field.

The breast tissues employed in this paper and their mean relative complex permittivity values at 1.1GHz [9] are listed in Table I. The coupling medium is a 80:20 glycerin:water liquid with complex permittivity $23.3 - j18.46$ [10] that mimics the average constitutive parameters of the breast.

The stochastic variation of the permittivity vector $\boldsymbol{\varepsilon}$ is accomplished through a set of m uncorrelated RVs $\boldsymbol{X} = (x_1, \ldots, x_m)$. In $\boldsymbol{\varepsilon}(\boldsymbol{X})$, only the permittivity values $\epsilon_n^{\mathrm{re}}, \epsilon_n^{\mathrm{im}}$ of cells with a center $\boldsymbol{r}_n$ inside the breast depend on $\boldsymbol{X}$, since the coupling medium is excluded from the random variation. We assume a standard normal distribution $\mathcal{N}(0,1)$ for each uncorrelated RV. As such, different realizations of $\boldsymbol{\varepsilon}$ are obtained by Gaussian sampling of $\boldsymbol{X}$. The MCV-LAR method is formulated for real variables, hence we employ the notation $E^{s,\mathrm{pa}}$, where pa stands for the real or imaginary part.

A. Monte-Carlo simulations

A straightforward method to extend the deterministic solver for stochastic modeling is the Monte-Carlo method. The exact computation of the scattered field E^s in the receiver locations is repeated for a large number N_{MC} of different permittivity vector realizations $\varepsilon(\boldsymbol{X})$. For a valid stochastic output distribution $E^{s,\mathrm{pa}}(\boldsymbol{X}, \boldsymbol{r}^{\boldsymbol{R}})$, at least 10000 samples are typically required, since the rate of convergence of MC is $\mathcal{O}(\sqrt{N_{MC}})$. We set $N_{MC} = 20000$ for all MC simulations performed with the deterministic solver.

TABLE I
RELATIVE COMPLEX PERMITTIVITY OF BREAST TISSUES AT 1.1GHZ

Tissue	skin	fat	gland	tumor
ϵ_r^{re}	35	12.6	32.7	53.4
ϵ_r^{im}	-23	-10.13	-20.92	-18.8

B. Polynomial chaos

As an alternative to a time-consuming Monte-Carlo analysis, the PC method provides a computational advantage when the number m of uncorrelated RVs $\boldsymbol{X} = (x_1, ..., x_m)$ is sufficiently small. The scattered field $E^{s,\mathrm{pa}}(\boldsymbol{X})$ at a given receiver position is expressed as an infinite expansion of orthogonal polynomials $\phi_k(\boldsymbol{X})$ of the uncorrelated RVs. The polynomials have to form a complete set and are selected to yield optimal convergence. To make the method computationally tractable, the expansion is truncated to $K+1$ terms,

$$E^{s,\mathrm{pa}}(\boldsymbol{r}^{\boldsymbol{R}}, \boldsymbol{X}) = \sum_{k=0}^{K} c_k^{\mathrm{pa}}(\boldsymbol{r}^{\boldsymbol{R}}) \phi_k(\boldsymbol{X}), \quad (2)$$

where the coefficients c_k^{pa} are real numbers and the dependence of $E^{s,\mathrm{pa}}$ on $\boldsymbol{X}$ is fully captured by the polynomials. In the following, the dependencies on $\boldsymbol{r}^{\boldsymbol{R}}$ are omitted. According to the Wiener-Askey scheme [11], the optimal polynomials for standard normally distributed input RVs are products of univariate probabilists' Hermite polynomials,

$$\phi_k(\boldsymbol{X}) = \prod_{i=1}^{m} \phi_{k_i}(x_i) = \prod_{i=1}^{m} H_{k_i}(x_i), \quad (3)$$

with H_{k_i} the probabilists' Hermite polynomial of order k_i related to the i^{th} uncorrelated RV. Let p denote the maximum degree that is allowed for the product of the polynomials in (2). It follows that for each term $\sum_{i=1}^{m} k_i \leq p$. For given values of m and p, the total number of terms in (2) then is $K+1 = \frac{(p+m)!}{p!m!}$, which grows extremely fast with m and p.

C. Cramér-von Mises criterion

To assess whether two output distributions, in the form of cumulative distribution functions (CDFs), are to be considered identical or different, we use the Cramér-von Mises (CVM) criterion [5]. The null hypothesis of the CVM criterion states that two arrays of values stem from the same distribution. The test of significance assesses the strength of the evidence against the null hypothesis. Thereto, we compute T_{norm}, i.e., the normalized T-value for the case of two large arrays. The value of T_{norm} corresponding to a specific significance level α is denoted by $t_{norm,\alpha}$. Consider for example a significance level of $\alpha = 5\%$. The value $t_{norm,5\%} = 0.461$ is chosen as the threshold for the CVM criterion. Under the null hypothesis, the probability of obtaining a T_{norm} value larger than or equal to $t_{norm,5\%}$ is 5%. If the T_{norm} value calculated from both arrays is smaller than 0.461, the null hypothesis is accepted and the distributions are considered equal, otherwise it is rejected.

D. Modified cross-validation least-angle regression

The MCV-LAR method yields a sparse expansion by iteratively selecting the polynomials ϕ_k that affect the output values most [4]. The computational effort of the method depends on the number of candidate polynomials. To reduce the number of candidate polynomials, a hyperbolic truncation scheme is applied. The polynomials ϕ_k thus have to satisfy the condition $(\sum_{i=1}^{m} k_i^Q)^{1/Q} \leq p$, with $k = (k_1, ..., k_m)$ the multi-index of ϕ_k, Q the Q-quasi norm and p a maximum value used in the truncation. The selected indices k are collected in the set denoted by $\mathcal{A}_Q^{m,p}$ with cardinality K'. The total number of terms in the HTPC expansion is denoted by $K'+1$. The MCV-LAR method is based on L deterministic model evaluations. Firstly, the LAR algorithm is applied to select the predictors, i.e., the polynomials in the HTPC expansion with a non-zero coefficient. The HTPC coefficients c_k are all initialized to 0 and the current residual is equal to the vector of L deterministic model evaluations denoted by $\boldsymbol{Y}$. In every iteration t, the polynomial that is most correlated with the current residual is transferred from $[\mathcal{A}^{(t)}]^c$ to $\mathcal{A}^{(t)}$, with $[\mathcal{A}^{(t)}]^c = \mathcal{A}_Q^{m,p} \backslash \mathcal{A}^{(t)}$, where $[\mathcal{A}^{(t)}]^c$ and $\mathcal{A}^{(t)}$ denote the set of candidate and active indices, respectively. At the start ($t = 0$), the active set is empty. After t iterations, $(K'+1-t)$ indices remain in the candidate set. At the start of the t^{th} iteration, $[K'+1-(t-1)]$ indices remain in the candidate set and the current residual is equal to $\boldsymbol{Y} - \boldsymbol{Y}^{t-1}$. The vector $\boldsymbol{Y}^{t-1}$ represents the predicted values in the L data points based on the metamodel constructed by the $(t-1)$ active polynomials and non-zero coefficients. The correlations of the $[K'+1-(t-1)]$ candidate polynomials with the current residual $\boldsymbol{Y} - \boldsymbol{Y}^{t-1}$ are computed, after which the polynomial with the highest correlation is selected. Next, the coefficients have to be updated. They are moved towards the least-square coefficients of the active predictors on the current residual until the new polynomial from $[\mathcal{A}^{(t)}]^c$ has an equal correlation with the current residual as the active set has. The two steps are repeated until $t = t_{max} = \min(K'+1, L)$. Secondly, the MCV method is applied to the t_{max} metamodels delivered by the LAR method. Their coefficients are recomputed with ordinary least-squares regression. The metamodel corresponding to the lowest value of the corrected Leave-One-Out error is retained and constitutes the stochastic solution for $E^{s,\text{pa}}$.

III. Results

A schematic breast tissue profile that is representative for the profiles in this paper is shown in Fig. 1. Its diameter is 0.12m and it comprises the four tissues listed in Table I: a thin layer of skin on a concentric layer of fat, containing a tumor with diameter 0.02m positioned at the bottom nearby the source, and in the center a circular region made up of randomly ordered cells of gland and fat. The square domain is discretized into $N = 257 \times 257 = 66049$ cells with size $\Delta = 0.000547$m, hence the number of cells inside the breast equals 37825. We verified by means of analytical solutions that the maximum discretization error over a wide range of tissue properties for cylinders of this size is $3 \cdot 10^{-3}$.

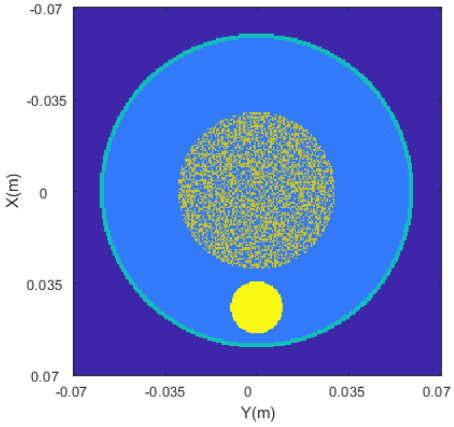

Fig. 1. Breast tissue profile.

In a first step we optimize the parameters of the HTPC MCV-LAR method in terms of the trade-off between accuracy and computational speed. This means selecting adequate values for Q and p in the HTPC expansion and for the number L of deterministic evaluations of $E^{s,\text{pa}}$ needed by the MCV-LAR method. The value of Q influences the number of candidate indices $K'+1$ allowed in the HTPC expansion and thereby the computational cost, which is of the order $\mathcal{O}(L(K'+1)^2 + (K'+1)^3))$ for the LAR method. We set $Q = 0.4$. The values of p and L are optimized for the profile of Fig. 1. Next we validate the optimized method on additional profiles and apply it to the detection of a tumor.

The optimization is achieved by comparing the output distributions of candidate sparse HTPC expansions as a function of the parameter under investigation with the output distribution from Monte-Carlo simulations performed with the deterministic forward solver. Each output distribution is generated with 20000 input samples $\boldsymbol{X}$. The CVM criterion with $\alpha = 1\%$ is used to decide whether both arrays—each of 20000 CDF values—stem from the same distribution. An accurate sparse HTPC expansion then corresponds to a T_{norm} value below $t_{norm,1\%} = 0.743$.

A. Profiles with a limited number of uncorrelated RVs

We first consider a simple stochastic permittivity model by letting all breast cells in the grid that belong to one tissue adopt the same permittivity value $\varepsilon_{tissue}^{\text{pa}} = \mu_{tissue}^{\text{pa}} + \sigma_{tissue}^{\text{pa}} x_{tissue}^{\text{pa}}$, where μ_{tissue}^{pa} and $\sigma_{tissue}^{\text{pa}}$ are the mean value (Table I) and the standard deviation, respectively, and where x_{tissue}^{pa} is one of the $m = 8$ uncorrelated RVs in $\boldsymbol{X}$, with m equal to twice the number of tissues present in the breast. The variance σ^2 is set to 10% of the mean value μ, hence

$$\begin{aligned}
\epsilon_{\text{skin}}^{\text{re}} &= 35 + 1.87x_1 & \epsilon_{\text{skin}}^{\text{im}} &= -23 + 1.52x_2 \\
\epsilon_{\text{fat}}^{\text{re}} &= 12.6 + 1.12x_3 & \epsilon_{\text{fat}}^{\text{im}} &= -10.13 + 1.01x_4 \\
\epsilon_{\text{gland}}^{\text{re}} &= 32.7 + 1.81x_5 & \epsilon_{\text{gland}}^{\text{im}} &= -20.92 + 1.45x_6 \\
\epsilon_{\text{tumor}}^{\text{re}} &= 53.4 + 2.31x_7 & \epsilon_{\text{tumor}}^{\text{im}} &= -18.8 + 1.37x_8 \quad (4)
\end{aligned}$$

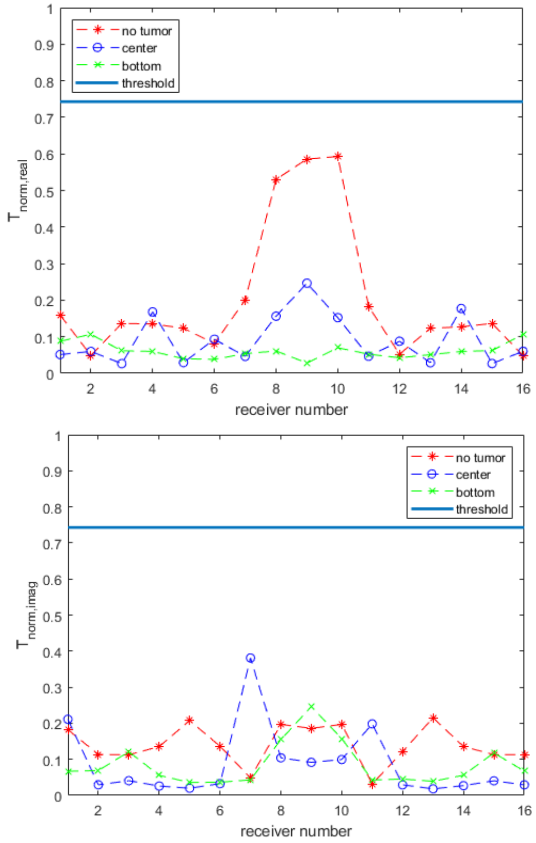

Fig. 2. Validation of the optimized parameters: T_{norm} for $E^{s,\mathrm{re}}$ (upper) and $E^{s,\mathrm{im}}$ (lower), for profiles without a tumor (red crosses) and with a tumor in the center (blue circles). The horizontal line is at $t_{norm,1\%}$.

To optimize the value of p, sparse HTPC expansions $E^{s,\mathrm{pa}}(\boldsymbol{X})$ are computed for a range of p-values between 5 and 15 for $L = 250, 500, 750$. For each expansion, an output CDF is generated, as mentioned above. The maximum values over all receivers of T_{norm} appear to be smallest (lower than 0.4) for $p = 15$. The computation time for constructing the HTPC expansions is nearly proportional to L and is nearly independent of p. This is due to the low number of uncorrelated RVs m, hence of card$(\mathcal{A}_Q^{m,p})$, which implies that the computation time of the LAR method is negligible compared with the time needed for the L deterministic evaluations. Consequently selecting $p = 15$ yields the most accurate approximation and has no negative impact on the computation time. Next, sparse HTPC expansions with $p = 15$ are computed for a range of L-values between 50 and 500. The maximum values over all receivers of T_{norm} reach a minimum value of $0.26 < t_{norm,1\%}$ at $L = 250$. We thus set $Q = 0.4$, $p = 15$ and $L = 250$ as optimal parameters for the the following.

The optimized HTPC MCV-LAR method is validated on two additional profiles. Starting from Fig. 1, one profile is obtained by removing the tumor and another one by moving the tumor to the center into the glandular region. Fig. 2 shows T_{norm} at the different receivers, obtained by applying the CVM criterion to the output distributions from the sparse HTPC expansions and the Monte-Carlo simulations with the deterministic solver. The $t_{norm,1\%}$ threshold is never exceeded, hence the sparse HTPC expansions can be considered as accurate. It was observed that the HTPC MCV-LAR method was approximately 50 times faster than Monte-Carlo sampling. For larger values of m, the computation time increases significantly.

We finally discuss tumor detectability for two scenarios: a tumor with diameter d is located either at the bottom of the breast near the source (Fig. 1), or in the center of the breast. The diameter of the tumor is varied from $d = 0.02$m down to $d = 0.005$m in three equal steps, while the position of its center is not altered. The number of receivers is increased to 160 to improve the resolution of the plots.

Tumor detectability is studied by comparing the profiles with and without a tumor in two ways: stochastically and deterministically. We now choose to expand the amplitude $|E^s|$, since it is experimentally harder to accurately measure the phase. On the one hand, $|E^{s,\mathrm{tumor}}|$ and $|E^{s,\mathrm{notumor}}|$ are expanded with the optimized HTPC MCV-LAR method. Both output distributions are compared with each other using the CVM criterion. The resulting values for T_{norm} are shown in the upper figures of Figs. 3 and 4 for the tumor in the center and at the bottom, respectively. Note that a significance level is not indicated, due to a lack of a relation between T_{norm} and the discretization error. On the other hand, $|E^{s,\mathrm{tumor}}|$ and $|E^{s,\mathrm{notumor}}|$ are computed with the deterministic solver for the mean permittivity values, i.e., with $\boldsymbol{X} = (x_1, ..., x_m) = \mathbf{0}$. The relative amplitude difference for the problem with and without a tumor is defined by

$$\Delta_{\mathrm{AMPL},0} = \left| \frac{|E^{s,\mathrm{tumor}}(\mathbf{0})| - |E^{s,\mathrm{notumor}}(\mathbf{0})|}{|E^{s,\mathrm{notumor}}(\mathbf{0})|} \right|. \quad (5)$$

The values of $\Delta_{\mathrm{AMPL},0}$ are shown in the lower figures of Figs. 3 and 4. Detection of the tumor is only possible if they exceed the discretization error of $3 \cdot 10^{-3}$, indicated by the horizontal line.

Firstly, consider the scenario with the tumor at the bottom near the source. In Fig. 4 the tumor is detectable at almost every receiver for all considered tumor sizes. It is more detectable at the receivers close to it. When it is not detectable at a certain receiver, that receiver is located at the other side of the breast with respect to the tumor and source and its exact location depends on the size of the tumor. Next, consider the scenario with the tumor at the center. In Fig. 3 the values of T_{norm} and the number of receivers where $\Delta_{AMPL,0}$ is larger than the discretization error decrease with decreasing size d, but even the smallest tumor with size $d = 0.005$m is detectable at some receivers further away from the source; it is the least detectable at the receivers close to the source. In reality, when trying to detect a tumor, every antenna location once serves as a source location. Tumor tissue at the center of the breast thus may be the hardest to detect, since the wave has to travel the longest distance.

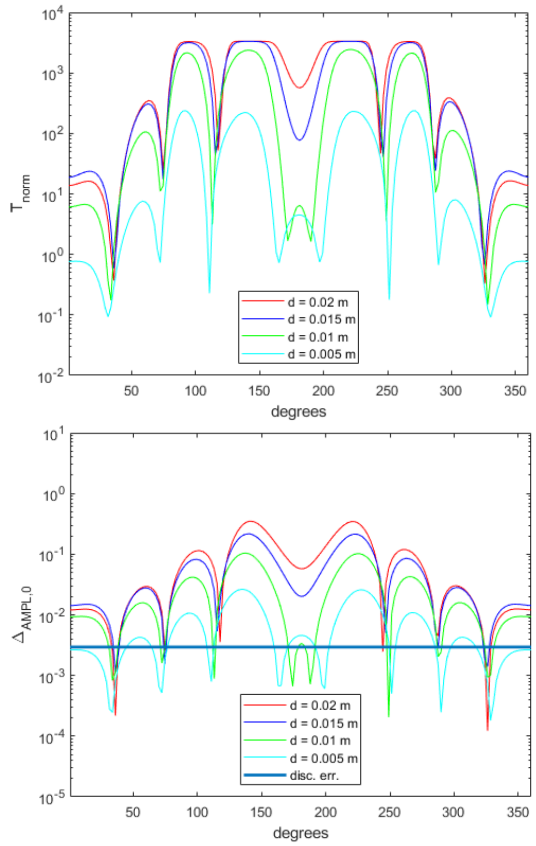

Fig. 3. Detectability of a tumor in the center, with tumor sizes 0.02m, 0.015m, 0.01m and 0.005m: T_{norm} based on $|E^{s,\mathrm{tumor}}|$ and $|E^{s,\mathrm{notumor}}|$ (upper) and $\Delta_{\mathrm{AMPL},0}$ (lower) as a function of the receiver position. The horizontal line indicates the discretization error.

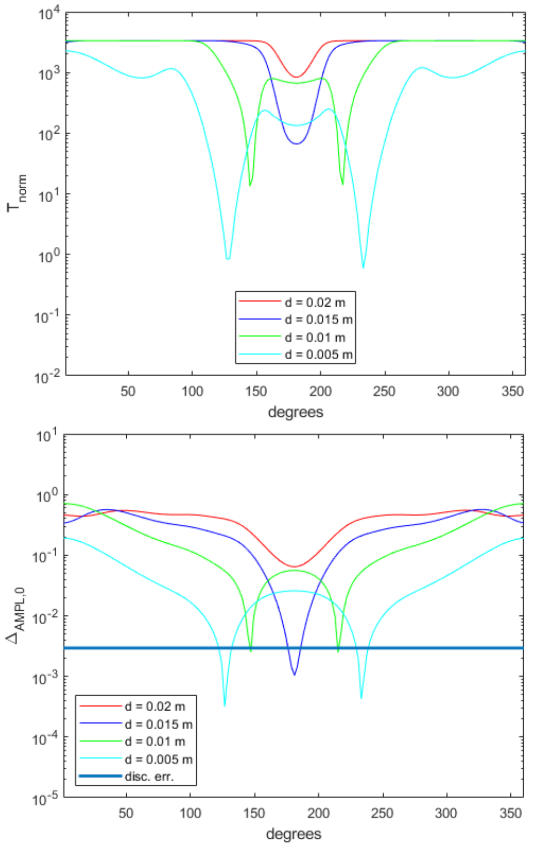

Fig. 4. Detectability of a tumor at the bottom, with tumor sizes 0.02m, 0.015m, 0.01m and 0.005m: T_{norm} based on $|E^{s,\mathrm{tumor}}|$ and $|E^{s,\mathrm{notumor}}|$ (upper) and $\Delta_{\mathrm{AMPL},0}$ (lower) as a function of the receiver position. The horizontal line indicates the discretization error.

B. Profiles with a larger number of uncorrelated RVs

We now consider a more realistic stochastic permittivity model by letting each grid cell inside the breast take a different complex permittivity value, represented by two RVs, one for the real and one for the imaginary part. For the profile of Fig. 1 with 37825 cells inside the breast, there are a total of $M = 75650$ RVs. In reality, the permittivity values of cells that belong to a same tissue are spatially correlated. The behavior of their RVs is characterized by their spatial correlation profile, which we assume to be Gaussian. The covariance matrix is defined as

$$\boldsymbol{\Sigma}_{ij} = \sigma_i \sigma_j \exp\left(\frac{-|\boldsymbol{r}_i - \boldsymbol{r}_j|^2}{L_c^2}\right), \tag{6}$$

with $i, j = 1, ..., M$, where M denotes the total number of correlated RVs, where σ_i and σ_j are the standard deviations and $\boldsymbol{r}_i$ and $\boldsymbol{r}_j$ the position vectors of the grid cells corresponding to the i^{th} and j^{th} correlated RV, respectively, and where L_c is the correlation length. The correlation length depends on the type of tissue and is chosen as $L_c = \lambda_{tissue}/5$, with λ_{tissue} the wavelength in the tissue. There is no correlation between different tissue types and the variances are as previously.

A PC expansion is based on m uncorrelated RVs and the run-time of the MCV-LAR method is dependent on m. The aim is to reduce the number M of correlated RVs significantly without loss of vital information. To reduce the dimensionality of the stochastic behavior, we apply the Karhunen-Loéve Transform (KLT), which is based on an eigenvalue decomposition of the covariance matrix

$$\boldsymbol{\varepsilon}^T = \boldsymbol{\mu}^T + \boldsymbol{U}\boldsymbol{\Lambda}^{1/2}\boldsymbol{X}^T \tag{7}$$

where T stands for transpose, $\boldsymbol{\mu}$ is the mean of the permittivity vector $\boldsymbol{\varepsilon}$, $\boldsymbol{\Lambda}$ is the $m \times m$ eigenvalue matrix corresponding to the m largest eigenvalues, $\boldsymbol{U}$ is the $M \times m$ eigenvector matrix and $\boldsymbol{X}$ denotes the vector of m uncorrelated RVs retained by KLT. The uncorrelated RVs corresponding to the largest eigenvalues contain the highest amount of information.

Considering the profile of Fig. 1, applying an eigenvalue decomposition to a 75650×75650 covariance matrix $\boldsymbol{\Sigma}$ requires an excessive amount of memory. Since with the chosen cell size $\Delta = 0.000547$m the correlation between the correlated RVs of neighboring cells is very strong, it is possible to approximate the distributions of 75650 correlated RVs by 18898 correlated RVs, i.e., on a coarser grid with a cell size $\Delta = 0.001094$m, without significantly influencing the approximations of the scattered field distributions at the receivers. Consequently, the KLT is performed on the smaller covariance matrix and afterwards, the matrix $\boldsymbol{U}\boldsymbol{\Lambda}^{1/2}$

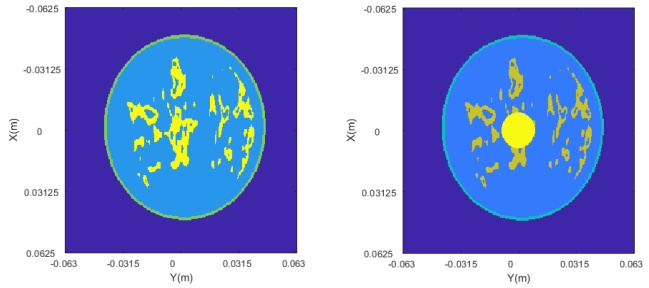

Fig. 5. ACR class 2 breast profile without (left) and with (right) a tumor added.

is extended again to $M = 75650$ rows, in order not to increase the discretization error.

The optimization of the HTPC MCV-LAR parameters p, Q, L and now also m is conducted in a similar way to that in the previous examples. The optimal settings are found to be $Q = 0.4$, $p = 6$ and $L = 1300$, while for m the values depend on the location of the receiver: $m = 100$ for receivers nearby the source and $m = 50$ for receivers opposite to the source.

We validate the optimized HTPC MCV-LAR method for a more realistic breast profile, that is based on a ACR class 2 (scattered fibroglandular) numerical breast phantom from the UWCEM database [12]. A coronal slice is shown in the left figure of Fig. 5. The originally different types of glandular tissues and the transitional tissue are grouped together into one single tissue gland and the different types of fatty tissues into one single tissue fat and the permittivity values from Table I are adopted. We construct a sparse HTPC expansion and perform Monte-Carlo simulations with the deterministic solver for $m = 1000$ uncorrelated RVs. Both output distributions, represented by 20000 samples each, are compared with the CVM criterion. For all receivers, the resulting values of T_{norm} are below $t_{norm,1\%}$.

We now examine tumor detectability for two different profiles: starting from the profile in the left figure of Fig. 5 a tumor with diameter $d = 0.02$m is positioned either in the center (right figure of Fig. 5), or at the bottom near the source. The scattered fields $|E^{s,\text{tumor}}|$ and $|E^{s,\text{notumor}}|$ are expanded with the optimized HTPC MCV-LAR method. Both output distributions are compared with each other using the CVM criterion. The values of T_{norm} are shown in Fig. 6. Similar to the results from the previous examples, the tumor located near the source is the most easily detectable.

Conclusion

Sparse hyperbolically truncated polynomial chaos expansions of the electric field scattered by a breast with random variations in its complex permittivity profile were constructed with the modified cross-validation least-angle regression method, which required significantly less deterministic model evaluations than Monte-Carlo sampling. The method was optimized and validated for breast permittivity profiles involving a limited as well as a large number of random

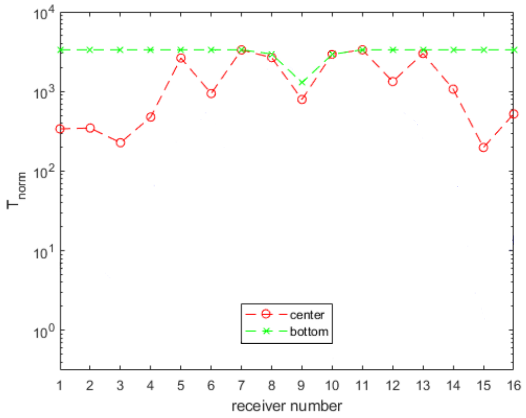

Fig. 6. Tumor detectability for ACR class 2 profiles: T_{norm} based on $E^{s,\text{tumor}}$ and $E^{s,\text{notumor}}$ for a tumor with size 0.02m in the center (red circles) and at the bottom (green crosses), as a function of the receiver position.

variables. In the latter case the Karhunen-Loéve transform was applied to reduce the stochastic dimensionality. Tumor detectability via the Cramér-von Mises criterion was illustrated for tumor sizes down to 0.005m.

References

[1] S. Kwon and S. Lee, "Recent advances in microwave imaging for breast cancer detection," *Int. J. Biomedical Imaging*, vol. 2016, Article ID 5054912, 26 p, 2016. https://doi.org/10.1155/2016/5054912.

[2] F. Bai, A. Franchois and A Pizurica, "3D microwave tomography with Huber regularization applied to realistic numerical breast phantoms," *Progress in Electromagnetics Research*, vol. 155 , pp.75-91, Jan. 2016.

[3] F. Boeykens, H. Rogier and L. Vallozzi, "An efficient technique based on polynomial chaos to model the uncertainty in the resonance frequency of textile antennas due to bending," *IEEE Trans. Antennas Propag.*, vol. 62, no. 3, pp. 1253–1260, March 2014.

[4] G. Blatman and B. Sudret, "Adaptive sparse polynomial chaos expansion based on least angle regression," *J. Comput. Phys.*, vol. 230, no. 6, pp. 2345–2367, March 2011.

[5] T. W. Anderson, "On the distribution of the two-sample Cramér-von Mises criterion," *The Annals of Mathematical Statistics*, vol. 33, no. 3, pp. 1148–1159, Sept. 1962.

[6] P. M. Meaney, M. W. Fanning, D. Li, S. P. Poplack and K. D. Paulsen,, "A clinical prototype for active microwave imaging of the breast," *IEEE Trans. Microw. Theory Tech.*, vol. 48, no. 11, pp. 1841–1853, Nov. 2000.

[7] J. H. Richmond, "Scattering by a dielectric cylinder of arbitrary cross section shape," *IEEE Trans. Antennas Propag.*, vol. 3, no. 3, pp. 334–341, May 1965.

[8] H. van der Vorst, "Bi-CGSTAB: a fast an smoothly convergent variant of Bi-CG for the solution of nonsymmetric linear systems," *SIAM J. Sci. Statist Comput.*, vol. 13, no. 2, pp. 631–644, March 1992.

[9] P. Mojabi and J. LoVetri, "Composite tissue-type and probability image for ultrasound and microwave tomography," *IEEE J. Multiscale Multiphys. Comput. Tech.*, vol. 1, pp. 26–35, April 2016.

[10] T. Rubaek, P. M. Meaney, P. Meincke and K. D. Paulsen, "Nonlinear microwave imaging for breast-cancer screening using Gauss-Newton's method and the CGLS inversion algorithm," *IEEE Trans. Antennas Propag.*, vol. 55, no. 8, pp. 2320–2331, Aug. 2007.

[11] D. Xiu and G. E. Karniadakis, "The Wiener-Askey polynomial chaos for stochastic differential equations," *SIAM J. Sci. Comput.*, vol. 24, no. 2, pp. 619–644, 2002.

[12] E. Zastrow, S. K. Davis, M. Lazebnik, F. Kelcz, B. D. Van Veen, and S. C. Hagness, "Development of anatomically realistic numerical breast phantoms with accurate dielectric properties for modeling microwave interactions with the human breast," *IEEE Trans. Biomed. Eng.*, vol.55, no.12, pp. 2792–2800, Dec. 2008.