

MICROWAVE CHARACTERIZATION OF DIELECTRIC MATERIALS USING BAYESIAN NEURAL NETWORKS

H. Acikgoz, Y. L. Bihan, O. Meyer, and L. Pichon

Laboratoire de Génie Electrique de Paris
CNRS UMR8507, SUPELEC
UPMC Univ. Paris 06, Univ. Paris-Sud
11 rue Joliot-Curie, Plateau de Moulon
91192 Gif-sur-Yvette Cedex, France

Abstract—This paper shows the efficiency of neural networks (NN), coupled with the finite element method (FEM), to evaluate the broadband properties of dielectric materials. A characterization protocol is built to characterize dielectric materials and NN are used in order to provide the estimated permittivity. The FEM is used to create the data set required to train the NN. A method based on Bayesian regularization ensures a good generalization capability of the NN. It is shown that NN can determine the permittivity of materials with a high accuracy and that the Bayesian regularization greatly simplifies their implementation.

1. INTRODUCTION

The determination of dielectric constant ε' and loss factor ε'' of dielectric materials using a measurement cell can be a difficult problem. For a coaxial open-ended probe, the relationship between the admittance measured at the discontinuity plane of the waveguide and the complex permittivity (direct model) can rarely be inverted. A solution is then to use a direct model in an iterative procedure aiming to reduce the difference between the measured observation (e.g., reflection coefficient or admittance) and the calculated one [1]. However, in many cases an analytical solution does not exist and numerical solutions are computationally expensive when used in an iterative procedure.

In order to avoid the drawbacks of an iterative procedure, a more efficient approach is to use a parametric model fitted thanks to a database constituted of examples of the relation linking the observation

to the permittivity. A good candidate are the neural networks (NN) since they can be universal and parsimonious approximators and can allow to approximate a wide range of functions provided that they are previously trained with a consistent data set [2]. NN is a well recognized technique in the field of microwave for solving source reconstruction problems or optimization problems [3–7]

The training aims to adjust the parameters of the NN so that it correctly approximates the physical behaviour of the system [8]. Whenever one likes to perform the training of a NN, a high number of hidden neurons can lead to have an overfitting phenomenon. In order to avoid this problem, a method based on Bayesian regularization [9] is introduced. This method is quite fast and leads to a good generalization capability.

NN were used in [10] and [11] in order to characterize fluids. The implemented method consists in the training of the NN by using measurements made on several standard materials (fluids). This approach implies to have all these materials and to know their frequency dielectric behaviour. Such a technique can appear to be close to the reality but at the same time it can be expensive and difficult to implement because of the necessity to elaborate standard samples.

The approach proposed in this paper is based on the combination of NN and a finite element method (FEM). The FEM is a robust and versatile computational method that can simulate the physical behaviour of the measurement cell. It is used to provide the equivalent admittance of the cell containing the material to be characterized by carrying out the direct problem. Thus, the FEM permits to be unrestricted on specific materials and enlarges the training set according to the request.

The FEM provides the data set, called training set, required for the NN adjustment. A data set is constituted of input (complex admittance, frequency) and output (ε' , ε'') pairs. The validity of the proposed NN-based inversion method is assessed by using a characterization protocol developed in the laboratory which has a restricted analytic solution [12]. Inversion results obtained with NN (from 1 MHz to 1.8 GHz) are presented and compared with other results obtained using a protocol having an analytic solution and that are inverted with an iterative procedure [13].

2. MEASUREMENT SETUP AND NUMERICAL METHOD

The characterization cell, called SuperPol, consists in a junction between a coaxial waveguide and a circular waveguide which is filled

in an inhomogeneous way. This means that the material under test is located at the continuity of the inner conductor and that it is held up by a Teflon crown ($\epsilon' = 2.1$) (Fig. 1(a)). The whole device is connected to an impedance analyzer.

The interest of such a configuration is that it allows having a homogenous electrical field in all the material which is favourable in several applications such as microwave heating. Compared with the protocol SuperMit (Fig. 1(b)) which has a homogeneous filling (consisting of a junction between a coaxial waveguide and a circular waveguide filled by the dielectric material), this configuration may be used from low frequencies until only a few gigahertz using the coaxial waveguide GR900 (inner diameter = 6.2 mm, outer diameter = 14.28 mm).

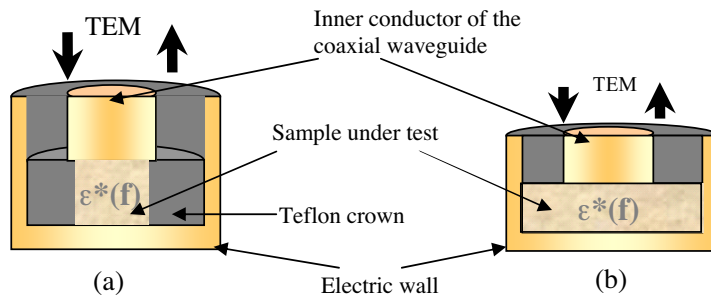


Figure 1. Measuring cells: (a) SuperPol. (b) SuperMit.

The electromagnetic problem is solved by using the FEM. The problem is expressed in terms of the electric field \mathbf{E} which satisfies the following harmonic wave equation:

$$\mathbf{curl} \left[-\frac{1}{i\omega\mu} \mathbf{curl}\mathbf{E} \right] - i\omega\epsilon\mathbf{E} = 0 \quad (1)$$

ω is the pulsation, ϵ ($\epsilon = \epsilon_0(\epsilon' - j\epsilon'')$) and μ are the permittivity and the permeability, respectively. ϵ_0 is the permittivity of the free space.

At the input of the coaxial waveguide, an excitation port is applied. An incident TEM electromagnetic field is prescribed. Perfect electric conductor (PEC) boundary conditions are applied on the surface of the waveguide conductors and at the end of the measurement cell. These boundary conditions are given by:

$$\mathbf{n} \wedge \mathbf{E} = 0 \text{ on conductors}$$

$$\mathbf{n} \wedge \mathbf{E} = \mathbf{n} \wedge \mathbf{E}_0 \text{ on the excitation plane with } E_0 \text{ a source field}$$

Databases are created using second order 3D tetrahedral vector finite elements. The degrees of freedom (DOFs) of the tangential vector elements are the projections of the electric field on the edges and faces of an element. The number of DOFs relevant to an element is 20. Thanks to the axial symmetry of the system, only an angular sector (5 degrees) of the geometry is meshed (Fig. 2). The system is meshed so that there are at least 10 elements per wavelengths. The whole mesh contains 4902 elements.

A perfect magnetic conductor (PMC) boundary is applied to both symmetry planes. This condition is given by:

$$\mathbf{n} \wedge \mathbf{H} = 0 \quad \text{On symmetry planes}$$

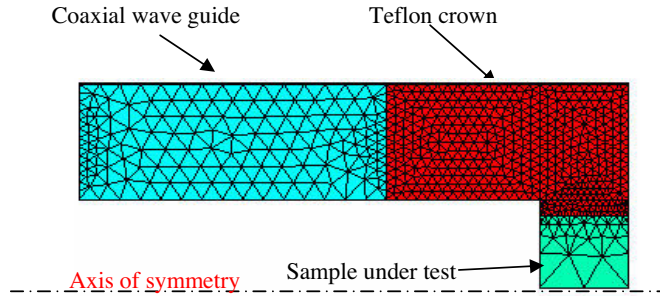


Figure 2. View of the meshed measurement cell.

3. IMPLEMENTATION OF THE NN

Two single-output networks corresponding to the two estimated quantities (ε' and ε'') are used. This approach is preferred to a single network (with two outputs) in order to avoid too complex NN including many internal parameters [14]. The inputs of a NN are the values of the complex admittance (real part G , imaginary part B) and the measurement frequency (f).

MLP (Multi-Layer Perceptron) NN are used. The retained structure consists in two layers NN with hyperbolic tangent activation functions in the hidden layer and an output layer constituted of a single neuron having a linear activation function. This structure of NN exhibits the properties of an universal and parsimonious approximator.

Figure 3 shows the structure of a NN with three inputs, N hidden neurons and an output.

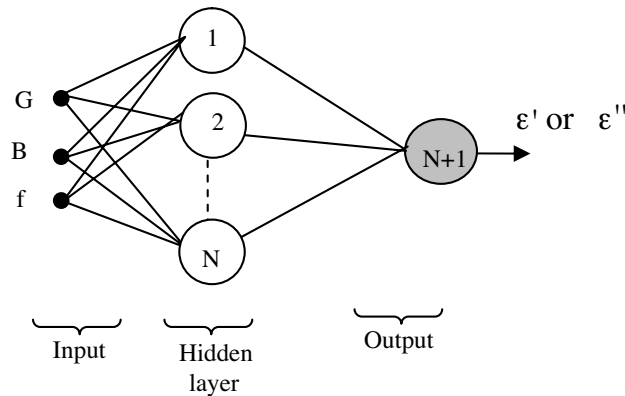


Figure 3. Structure of a NN.

Nevertheless, the design of the NN has to be achieved carefully. Indeed, in the classical approaches, a NN can be successfully used only if the number of hidden neuron is correctly chosen. As a matter of fact, a NN having too few neurons will not be able to learn correctly the training set data (underfitting). On the other hand, a NN having too many neurons can lead to an overfitting phenomenon: good learning of the training set data but poor generalization ability. That means that data not contained in the training set will be wrongly approximated. Consequently, in order to have a good learning and a good generalization, the NN designer has to determine the number of hidden neurons (and therefore the number of NN internal parameters). The most widely used method to define the size of a NN is the split sample procedure [15]. It consists in dividing the data set into two different sets: training and validation sets, different one from the other regarding the examples that they contain. The training of the NN (adjustment of its internal parameters) is carried out using the training set by iteratively minimizing an objective function, typically the mean square error (MSE), between the output of the NN and the one of the training set. Algorithms based on a first or second order expansion of the objective function are typically used. The validation set is used to determine the number of hidden neurons. This is done by calculating the MSE on the validation set for different NN having different number of hidden neurons. The optimal NN is the one which gives the lowest MSE on this set. It is obvious that this procedure is a consuming time task since it demands the creation of a validation set in addition to the training one and to train several NN of different sizes.

This paper presents a method based on the Bayesian regulariza-

tion that was first introduced in the field of NN by MacKay [9] to solve the problem of overfitting. It is based on the maximization of the a posteriori distribution of the NN parameters. Contrary to the previous method described above, this approach does require neither a validation set nor the systematic training of NN of different size. We present below an overview of the Bayesian regularization applied for fitting of the NN weights, i.e., the internal parameters of the NN.

Let us define D the training dataset. Assuming that the approximation error of these data is Gaussian, white and centred leads to the following distribution of the error:

$$P(D/w, \beta, M_i) = \frac{1}{Z_D(\beta)} \exp(-\beta E_D) \quad (2)$$

where w is the vector of the weights of M_i , the particular NN used. $E_D = \frac{1}{2} \sum_{N_D} (\varepsilon_{NN_i} - \varepsilon_i)^2$ where ε_{NN} , ε_i and N_D are the NN output, training set output and the number of training set examples, respectively. The summation is done over the training set examples. β is a parameter representative of the standard deviation of the noise and $Z_D = (2\pi/\beta)^{N_D/2}$.

The a priori distribution (i.e., without knowledge of the database) of the weights is also assumed as Gaussian, white and centred. It can be written as:

$$P(w/\alpha, M_i) = \frac{1}{Z_w(\alpha)} \exp(-\alpha E_w) \quad (3)$$

where $Z_W = (2\pi/\alpha)^{N_w/2}$, N_w is the number of NN weights and $E_w = \frac{1}{2} \sum_{N_w} w_j^2$. This assumption on the distribution is based on the knowledge that the weights can be either positive or negative and that it is reasonable to have small magnitude weights in order to obtain a smooth fitting allowing a good generalization capability.

The Bayesian regularization involves two levels of inference. Firstly, the weight parameters are estimated using Bayes' theory. The posterior probability density of the parameters w is:

$$P(w/D, \alpha, \beta, M_i) = \frac{P(D/w, \beta, M_i)P(w/\alpha, M_i)}{P(D/\alpha, \beta, M_i)} \quad (4)$$

By substituting (2) and (3) into (4), one obtains:

$$P(w/D, \alpha, \beta, M_i) = \frac{1}{Z_F(\alpha, \beta)} \exp(-(\beta E_D + \alpha E_w)) \quad (5)$$

where $Z_F = Z_w(\alpha)Z_D(\beta)P(D/\alpha, \beta, M_i)$.

Notice that, the optimal weights should maximize this posterior probability density. This is equivalent to minimizing the regularized objective function given by:

$$E_T = \beta E_D + \alpha E_w \quad (6)$$

The first term E_D is proportional to the mean square error function between the outputs of the training set and the ones of the NN. This is the usually used objective function for NN training. The second term E_w is a regularization term (weight decay term). A low value of the regularizer E_w , i.e., a low magnitude of the NN parameters, leads to a NN having an output varying smoothly and so thwarting the overfitting phenomenon.

In order to have a good generalization capability, the hyperparameters α and β have to be tuned. On one hand, if $\alpha \ll \beta$, E_D will be minimized but the generalization capability of the NN will not be ensured. On the other hand, if $\alpha \gg \beta$, the network will have a poor learning capability and will not learn the training set data. That is why, one must have to reach to a compromise as regard to α and β . This adjustment has also been performed by a Bayesian theory in a second level of the inference.

The objective function parameters α and β are optimized by applying the Bayes' rule:

$$P(\alpha, \beta/D, M_i) = \frac{P(D/\alpha, \beta, M_i)P(\alpha, \beta/M_i)}{P(D/M_i)} \quad (7)$$

If an uniform prior density $P(\alpha, \beta/M_i)$ is assigned to (α, β) , the maximization of the posterior density probability $P(\alpha, \beta/D, M_i)$ for these parameters is obtained by maximizing the evidence for α, β which can be written as:

$$P(D/\alpha, \beta, M_i) = \frac{Z_F(\alpha, \beta)}{Z_W(\alpha)Z_D(\beta)} \quad (8)$$

A Taylor expansion of $P(w/D, \alpha, \beta, M_i)$ around the most probable weight value of $w(w^{MP})$, which is obtained in the first level of inference by maximizing the posterior density probability for w , allows writing $Z_F = (2\pi)^{N_w/2} \det(((\mathbf{H}^{MP})^{-1}))^{1/2} \exp(E_T(\mathbf{w}^{MP}))$. \mathbf{H} is the Hessian matrix of the regularized objective function calculated for w^{MP} .

Solving Eq. (8) for the most probable values for α and β , we can obtain [16]:

$$\alpha^{MP} = \frac{\gamma}{2E_w(w^{MP})}, \quad \beta^{MP} = \frac{N_D - \gamma}{2E_D(w^{MP})}$$

The parameter $\gamma = N_w - 2\alpha^{MP}tr(\mathbf{H}^{MP})^{-1}$ is defined as the effective number of parameters of the NN. It is a measure of how many parameters are effectively used to reduce the error function and to have a good generalization capability.

Let us remind that using Bayesian regularization, in order to accomplish a good learning process, the number of neurons must be sufficiently high. Beyond this minimum number of hidden neurons, whatever it is (i.e., number of total parameters N_w), the effective number of parameter γ will remain less than N_w and therefore the output of the NN will have a smooth variation (good generalization).

The training of the NN is done by adjusting w^{MP} et α and β at each iteration.

4. INVERSION OF SIMULATED DATA

This section gives some inversion results on simulated data that compare the two methods described above. NN were trained with a training set of 3000 examples. The training domain is as follows: the estimated parameter (ε' and ε'') space is regularly discretized, between 1 to 30 for the real relative permittivity (ε') and 0 to 30 for the imaginary relative permittivity (ε''), whereas the frequency space is discretized according to a logarithmic law between 1 MHz and 1.8 GHz..

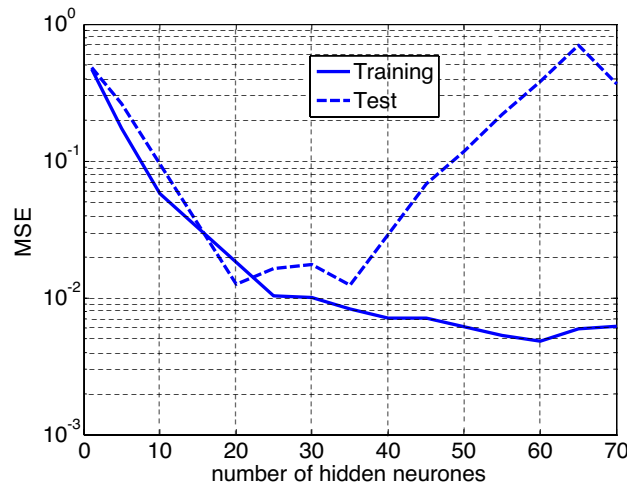


Figure 4. MSE obtained on training and test sets for different numbers of hidden neurones (real permittivity) with MSE objective function.

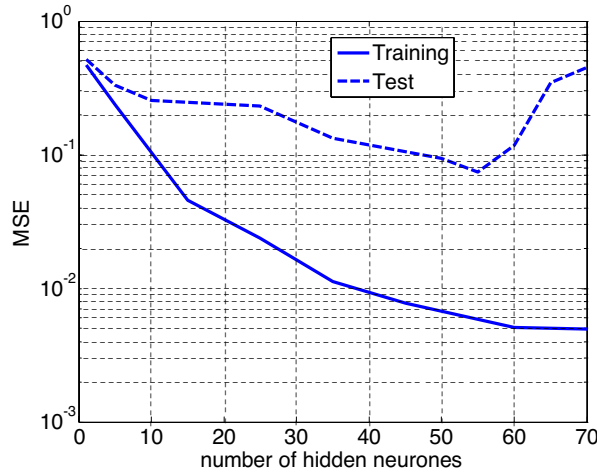


Figure 5. MSE obtained on training and test sets for different numbers of hidden neurones (imaginary permittivity) with MSE objective function.

The inversion performances are evaluated on a test set constituted of 500 randomly chosen examples in the training domain. Each example contains the complex admittance, the frequency and one of the two outputs (ε' or ε'').

Figures 4 and 5 show that, when using the MSE as objective function, an overfitting phenomenon effectively appears when the number of hidden neurones increases. This is underlined by the increase of the MSE calculated on the test set.

On the contrary, Figs. 6 and 7 show that the error obtained using the bayesian regularization on the training and test set vary very few beyond a sufficient number of neurones. The overfitting phenomenon occurring usually with NN does not appear in spite of the increase of the NN size.

These results show that NN inversion with Bayesian regularization is more suitable for training than “split sample” method. Consequently, in the following, Bayesian regularization will be applied to NN to determine the complex permittivity of a dielectric material. In the remainder, the results below have been obtained with NN that have 20 hidden neurones for each of ε' and ε'' .

Before doing the inversion of data issued from measurement, one checks the NN capability by comparing its output with the one contained in the test set. Fig. 8 shows that the outputs of the NN and the ones of the test set are in good agreement.

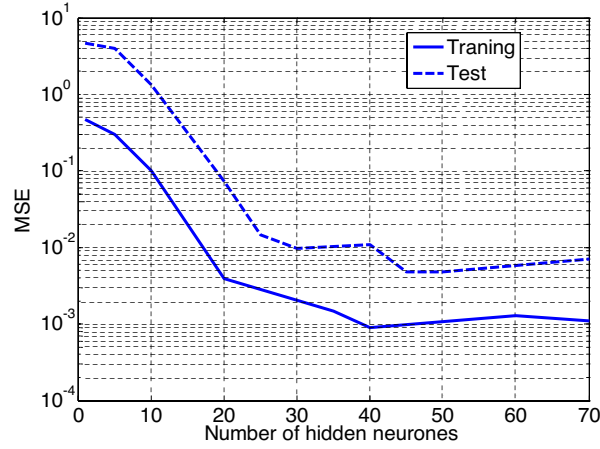


Figure 6. MSE obtained on the test set for different number of hidden neurons (real permittivity) with Bayesian regularization.

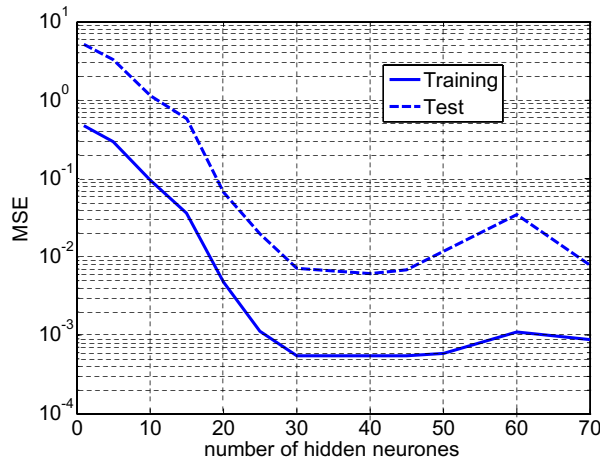


Figure 7. MSE obtained on the test set for different number of hidden neurons (imaginary permittivity) with Bayesian regularization.

5. EXPERIMENTAL RESULTS

Since the networks correctly approximate the outputs contained in the test set, experimental measurements can be presented at the input (G , B , and f) in order to evaluate the permittivity of the material under investigation on the desired frequency band.

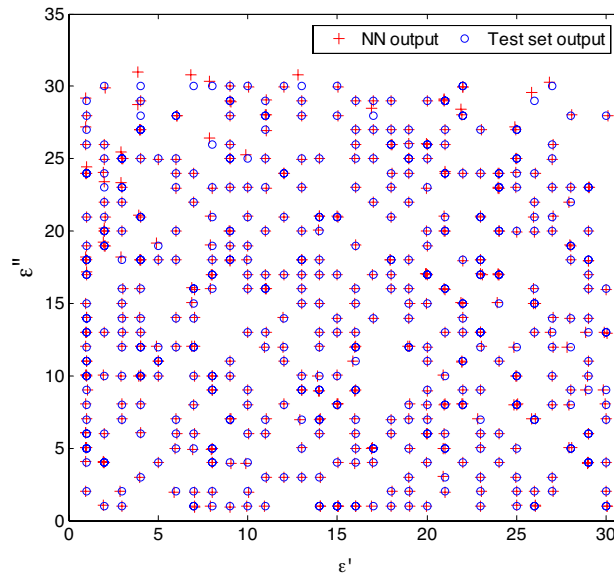


Figure 8. Comparison between NN output and the one contained in the test set.

Measurements have been carried out by using an impedance analyzer Agilent 4291A on an ethanol sample whose dielectric characteristics are known. The thickness of the sample under test is 2.9 mm. The coaxial guide is a standard GR900 (14 mm). The measurement frequency band is from 1 MHz to 1.8 GHz.

First of all, the permittivity of the material is determined using NN with 20 neurons (for both ϵ' and ϵ'') trained on the one hand with MSE objective function and on the other hand with Bayesian regularization. In this case, the generalization is quite good for both procedures as shown in Fig. 9. Indeed, this figure shows the good agreement between the results obtained with SuperPol with NN and those obtained with the iterative inversion based on SuperMit's. An analytical solution obtained by Debyes law is also represented.

The broad band permittivity of the ethanol obtained by Debye's law has also been represented. Results of the NN trained with Bayesian regularization show a good agreement with the analytical ones. The slight difference between the Debye's results and those obtained by the inversion of the experimental data (NN and iterative inversion) is due to the experimental conditions such as the temperature, the presence of impurities in the sample, etc. . . . Furthermore, Debye's model doesn't take into account the conduction phenomenon that appears at low

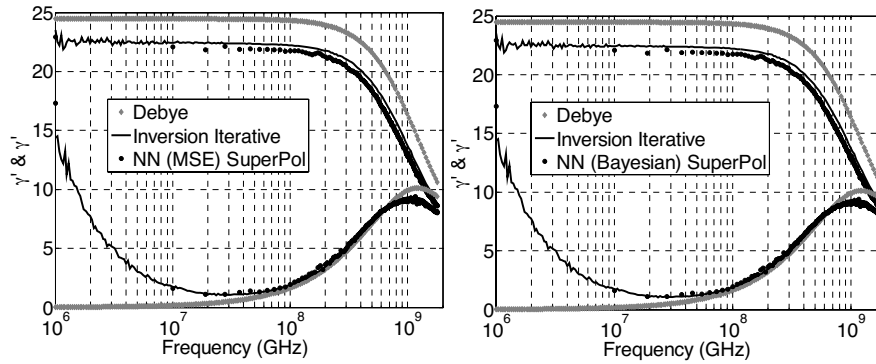


Figure 9. Permittivity of ethanol obtained with 20 neurons: MSE (left), Bayesian regularization (right).

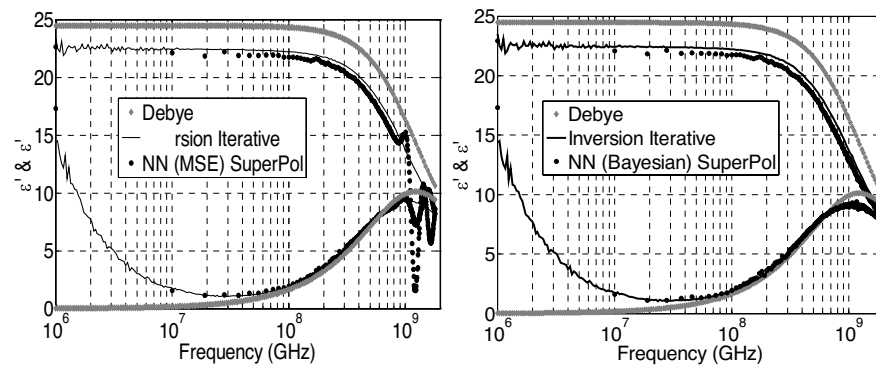


Figure 10. Permittivity of ethanol obtained with 50 hidden neurons: MSE (left), Bayesian regularization (right).

frequency on the imaginary part of the permittivity.

The generalization capability of the two previously described methods of NN learning has also been studied by chosen a high number of hidden neurons (and so of NN internal parameters). 50 hidden neurons in the hidden layer would be sufficient to remark the good generalization capability of the Bayesian regularization and on the contrary the bad generalization of the method with the MSE as the objective function. One can see on Fig. 10 that using MSE as objective function, NN badly generalise and this phenomenon mainly appears at high frequencies, whereas with the Bayesian regularization the permittivity of ethanol is correctly obtained even at high frequencies.

6. CONCLUSIONS

The interest of NN inversion in microwave dielectric characterization has been shown. MLP NN are good candidates to solve inverse problem that has no analytical solution. The Bayesian regularisation has been successfully applied and it is noticed that the NN generalize well when they are trained according to the Bayesian framework. The use of the Bayesian regularization allows reducing the criticity of the hidden neuron number. Compared to the split-sample procedure, the training became quite faster (reduction of about 80% of the training time) and less example data are required because of the unnecessary to have a validation set.

REFERENCES

1. Belhadj-Tahar, N. and A. Fourier-Lamer, "Broad-band analysis of discontinuity used for dielectric measurement," *IEEE MTT*, Vol. 34, 346–349, 1986.
2. Hornik, K., M. Stinchcombe, and H. White, "Multilayer feedforward networks are universal approximators," *Neural Networks*, Vol. 2, 359–366, 1989.
3. Yildiz, B. and M. Turkmen, "Quasi-static models based on artificial neural networks for calculating the characteristic parameters of multilayer cylindrical coplanar waveguide and strip line," *Progress In Electromagnetics Research B*, Vol. 3, 1–22, 2008.
4. Zainud-Deen, S. H., H. A. Malhat, K. H. Awadalla, and E. S. El-Hadad, "Direction of arrival and state of polarization estimation using radial basis function neural network (RBFNN)," *Progress In Electromagnetics Research B*, Vol. 2, 137–150, 2008.
5. Guney, K., C. Yildiz, S. Kaya, and M. Turkmen, "Artificial neural networks for calculating the characteristic impedance of air-suspended trapezoidal and rectangular-shaped microshield lines," *J. of Electromagn. Waves and Appl.*, Vol. 20, No. 9, 1161–1174, 2006.
6. Ayestar'an, R. G. and F. Las-Heras, "Near field to far field transformation using neural networks and source reconstruction," *J. of Electromagn. Waves and Appl.*, Vol. 20, No. 15, 2201–2213, 2006.
7. Acikgoz, H., Y. Le Bihan, O. Meyer, and L. Pichon, "Neural networks for broad-band evaluation of complex permittivity using a coaxial discontinuity," *Eur. Phys. J. Appl. Phys.*, Vol. 39, 197–201, 2007.

8. Sarle, W. S., "Neural Network FAQ, Part 2 of 7: Learning, periodic posting to the Usenet newsgroup comp.ai.neural-nets," Available by <ftp://ftp.sas.com/pub/Neural/FAQ.html>.
9. MacKay, D. J. C., "Bayesian methods for adaptive models," Thesis of California Institute of Technology, 1992.
10. Bartley, P. G., R. W. McClendon, and S. O. Nelson, "Permittivity determination by using an artificial neural network," *IEEE Instrum. and Meas. Tech. Conference*, 27–30, 1999.
11. Tuck, D. and S. Coad, "Neurocomputed model of open-circuited coaxial probes," *IEEE Microwave Guided Lett.*, Vol. 5, 105–107, 1995.
12. Meyer, O., "Instrumentation pour un contrôle de processus de reticulation sous micro-ondes par caractérisation large bande," Thesis of the University Pierre et Marie Curie, 1996.
13. Belhadj-Tahar, N. E., A. Fourier-Lamer, and H. De Chanterac, "Broad-band simultaneous measurement of complex permittivity and permeability using a coaxial discontinuity," *IEEE Transactions on Microwave Theory and Techniques*, Vol. 38, No. 1, January 1990.
14. Courtois, S. and R. Phan-Tan-Luu, "Neural networks applied to the choice of an experimental design," *EDP Sciences*, Vol. 26, 304–310, Wiley-VCH, 1998.
15. Sarle, W. S., "Neural Network FAQ, Part 3 of 7: Generalization, periodic posting to the Usenet newsgroup comp.ai.neural-nets," Available by <ftp://ftp.sas.com/pub/Neural/FAQ.html>.
16. Foresee, F. D. and M. T. Hagan, "Gauss-Newton approximation to Bayesian learning," *IEEE Trans. Neural Networks*, Vol. 3, 1930–1935, 1997.