Neurocomputing ∎ (∎∎∎∎) ∎∎∎–∎∎∎

Contents lists available at ScienceDirect

Neurocomputing

journal homepage: www.elsevier.com/locate/neucom

Wood moisture content prediction using feature selection techniques and a kernel method

Hela Daassi-Gnaba*, Yacine Oussar, Maria Merlan, Thierry Ditchi, Emmanuel Géron, Stéphane Holé

Laboratoire de Physique et d'Étude des Matériaux (LPEM) PSL Research University, ESPCI-ParisTech Sorbonne Universités, UPMC Univ Paris 06 CNRS, UMR 8213, 10 rue Vauquelin, 75231 Paris Cedex 05, France

ARTICLE INFO

Article history:
Received 11 January 2016
Received in revised form
15 July 2016
Accepted 1 September 2016
Communicated by Feiping Nie

Keywords:
Moisture content
Nonlinear regression
LS-SVM
Feature selection

ABSTRACT

Wood is a renewable, abundant bio-energy and environment friendly resource. Woody biomass Moisture Content (MC) is a key parameter for controlling the biofuel product qualities and properties. In this paper, we are interested in predicting MC from data. The input impedance of half-wave dipole antenna when buried in the wood pile varies according to the permittivity of wood. Hence, the measurement of reflection coefficient, that gives information about the input impedance, depends directly on the MC of wood. The relationship between the reflection coefficient measurements and the MC is studied. Based upon this relationship, MC predictive models that use machine learning techniques and feature selection methods are proposed. Numerical experiments using real world data show the relevance of the proposed approach that requires a limited computational power. Therefore, a real-time implementation for industrial processes is feasible.

1. Introduction

The world energy consumption has highly increased due to the industrial development. Global warming is one of the main problems in the 21st century. A solution consists in using renewable energies that are zero carbon footprint. Indeed, government policies have been applied to motivate the use of renewable energies in order to get the EU target of 20% of energy consumption from renewable sources by 2020 [1]. Wood is the main source of biomass used for heating and represents more than half of the world renewable energy consumption [2]. Wood fuel calorific value depends on the weight of dry wood within the sample. However, in real conditions the delivered sample is never dry. Therefore, woody biomass Moisture Content (MC) becomes the main parameter to determine the energy content and thus the price of the fuel [3,4]. It is interesting for the wood-to-energy industry to measure the MC in delivered containers. The standard method for MC determination of solid fuels in Europe consists in drying several samples in an oven [5]. This method is based on the weight loss determination after a drying process that usually lasts 24 hours. It is a time consuming method and does not provide an average value of MC of a whole truck of bio fuel. To contribute to the wood-to-energy industrial progress, a challenge for scientific research is to propose a reliable method for measuring the MC that is both representative and rapid.

There are several methods available for the determination of MC. They can be divided in two groups: direct methods and indirect methods.

- Direct methods are based on drying processes such as the standard method. They are time-consuming and can only be applied on small samples [6]. Recently, for wood drying process, MC prediction modeling that uses statistical learning methods was proposed. In 2009, an improved neural network model was presented in [7] for predicting lumber MC. During the same year, soft sensor model based on Least Squares Support Vector Machines (LS-SVM) was proposed to predict wood MC [8]. Temperature and humidity are the inputs and MC is the model output. Later, in 2012, Wen et al. [9] proposed a MC predictive model based on SVM, where temperature and equilibrium MC are the model inputs and MC is the model output.
- Indirect methods operate undestructively and rapidly. Therefore, they are more suitable for the wood-to-energy industry requirements [4]. These methods are based on electrical, optical, radiometric, thermal, hygrometric or acoustic technologies. Both Infrared (IR) and Near-Infrared (NIR) techniques are surface measurement methods that can be only applied on bio fuel convoyed on a belt [3]. These methods are independent of the density of the material since they are sensitive to the surface or more generally to the vicinity of the sensor. Microwaves, having a penetration

* Corresponding author.

http://dx.doi.org/10.1016/j.neucom.2016.09.005
0925-2312/

depth of 15 cm are suitable for conveyor applications. However, microwaves as well as capacitive or Time-Domain-Reflectometric (TDR) methods are influenced by the density of the material up to 5% to the total variation [4]. Radio-frequency technology (RF) is reviewed in literature as the most suitable method for measurements of large samples. Density, mass and temperature of the material affect the signals. However, they can be compensated [10]. The RF measurements are used to estimate the dielectric constant information of the woody biomass. The processing of data from the RF measurements can be improved by the implementation of black box models. The implementation of statistical learning methods for MC prediction was proposed in [11]. The measurement system consisted of two barrels: the upper barrel shields the antenna and the lower barrel contains the samples. The two barrels are connected during measurements. The barrels act as a waveguide.

Both direct and indirect methods show that the use of machine learning techniques is a promising domain for MC prediction in wood biomass. Other successful applications of these techniques in the field of wood science are reported in [12,13] for prediction of thermal conductivity and dielectric loss factor, respectively.

In the present study, an indirect method based on RF measurements obtained with an antenna that is fully buried into the samples of wood chips is proposed. When the antenna is buried inside the wood its input impedance varies according to the wood permittivity. Hence, the reflection coefficient S_{11} is measured. It gives information about the input impedance. The modeling approach we propose is implemented according to the following three steps:

- Several measurements of the reflection coefficient S_{11} in the frequency-domain of the same sample are performed to take into account the variations caused by the heterogeneity of the samples. These measurements differ by changing the arrangement of the wood chips.
- The available measurements are used for building reflection coefficient models to estimate S_{11} as a function of the frequency.
- The outputs of the reflection coefficient models are used as inputs of the MC model. Therefore, the MC model allows to predict MC from estimated values of S_{11} using feature selection methods and the LS-SVM modeling technique which is a kernel method.

The reflection coefficient and the MC models were designed and optimized to be implemented in industrial real time applications. This paper begins by presenting the experimental setup we built to measure the S_{11} of a half-wave length dipole, using a small-scale system for bulk measurements. The modeling methodology we propose for MC prediction is described. The feature selection methods, the LS-SVM technique with a sophisticated validation procedure and the full modeling methodology are then described in detail. Finally, results of numerical experiments conducted with two species of wood chips and a half-wave dipole antenna of two different lengths are presented and discussed to show the effectiveness of the proposed method.

2. Setup description and data acquisition

The literature gives different definitions of MC, in this study MC is calculated on the basis of the wet sample according to following equation,

$$\mathrm{MC}(\%) = \frac{m_{water}}{m_{wet}} \times 100 = \frac{m_{water}}{m_{water} + m_{dry}} \times 100 = \frac{m_{wet} - m_{dry}}{m_{wet}} \times 100 \quad (1)$$

where m_{water} is the mass of water within the wood, m_{dry} is the mass of dry wood and m_{wet} is the whole mass of wet wood. The MC of wood chips varies largely from around 20% to around 50%. The relative permittivity ε_r of wood strongly changes in the presence of water. Molecules of water generally present a random orientation. However, when an electric field is applied, the molecules orient themselves according to the polarity of the field [14]. The dielectric properties of a material result from its response under the influence of an electric field and can be expressed by its permittivity $\varepsilon = \varepsilon_0 \times \varepsilon_r$, where ε_0, the permittivity of the vacuum, is constant and only depends on the units used. ε_r, as presented in Eq. (2), is a complex number that depends on the angular frequency ω:

$$\varepsilon_r(\omega) = \varepsilon'_r(\omega) - j\varepsilon''_r(\omega) \quad (2)$$

In this paper, the dielectric constant ε'_r of water is around 80 [15]. In comparison, ε'_r of dry wood is around 1.7 and that of air is equal to 1. Thus, the global permittivity ε of the material under test composed by water, wood and air is highly sensitive to the presence of water. A half-wave dipole antenna has a pure real input impedance at both the resonance frequency f_r and odd harmonics. When the generator that feeds the antenna has a real output impedance equal to the antenna input impedance, the reflection coefficient S_{11} presents minimal values at these frequencies. The frequency f_r is related to the length of the antenna l and to the electric properties of the material according to the following relation:

$$f_r = \frac{c}{2l\sqrt{\varepsilon_r}} \quad (3)$$

where c is the light speed in vacuum.

Therefore, for a given dipole antenna of length l, the measured reflection coefficient S_{11} depends on the relative permittivity ε_r and subsequently on the Moisture Content MC.

As illustrated in Fig. 1 (right), minima of S_{11} are observed at the resonance frequency and at odd harmonics. Their positions depend on the MC value. Our method is based on the measurement of the reflection coefficient S_{11} of the half-wave dipole antenna within a sample of wood chips using a 8722ES Vector Network Analyzer from Agilent Technologies [16–18]. Fig. 1 (left) shows the laboratory-scale measurement system that was used.

We recall that wood chips are heterogeneous material. In order to take into account the variations caused by the heterogeneity of the samples, several measurements of the reflection coefficient S_{11} in the frequency domain must be carried out for the same sample. Hence, the necessity for changing the arrangement of chips between the measurements.

3. Combined model structure of moisture content prediction

This section is dedicated to a brief description of the modeling methodology proposed and summarized in the block diagram illustrated in Fig. 2.

In our context, S_{11} measurements obtained with an RF antenna that is fully buried into the samples of wood chips are used. As mentioned above, it needs to take into account possible variations of S_{11} caused by the arrangement of chips around the antenna. The modeling approach is implemented according to the following three steps: (i) For a given MC value, several measurements of the

Please cite this article as: H. Daassi-Gnaba, et al., Wood moisture content prediction using feature selection techniques and a kernel method, Neurocomputing (2016), http://dx.doi.org/10.1016/j.neucom.2016.09.005

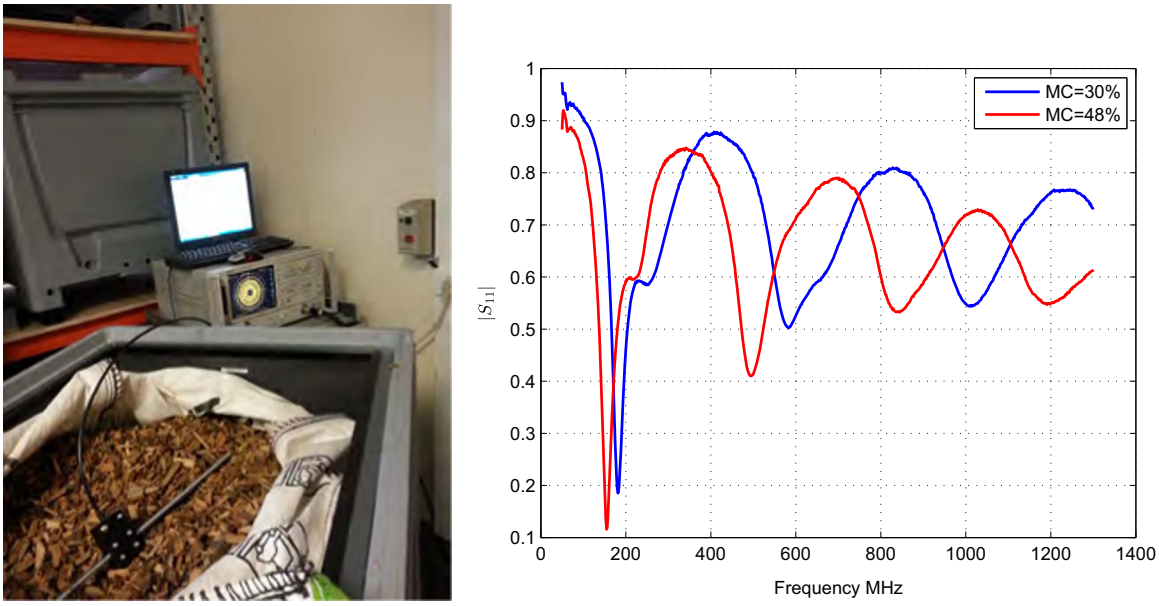


Fig. 1. The laboratory-scale measurement system (left). Modulus of S_{11} of the 54 cm length antenna for two MC values (right).

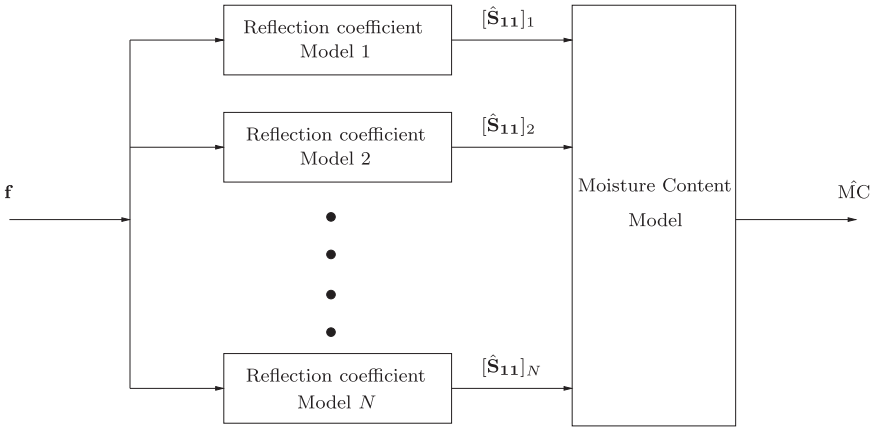


Fig. 2. Block diagram of the proposed combined model for MC prediction.

reflection coefficient S_{11} are performed in the frequency-domain which differ by the arrangement of the wood chips, (ii) using all the available measurements, reflection coefficient models are built to estimate S_{11} as a function of the frequency f, (iii) the outputs of the reflection coefficient models $\hat{\mathbf{S}}_{\mathbf{11}}$ are used as inputs of the MC model. Thereby, the MC model allows to predict MC from estimate values of S_{11}. The output of the MC model is $\hat{MC}$.

For N values of wood MC, the proposed modeling approach leads to a combined model structure: a set of N reflection coefficient models and a MC model.

The design of the reflection coefficient and the MC models requires the implementation of sophisticated methods for both variable selection and nonlinear black box models synthesis. Before going further in the design of both models methodology, a detailed description of the machine learning methods implemented in the study is given.

4. Variable ranking and selection methods

In modern modeling problems, it is often convenient to consider relevant variables and exclude irrelevant or redundant ones. The variable selection procedure consists in identifying the most representative subset of variables [19]. Variable selection methods were advantageously implemented to enhance data separation problems [20,21] or to inspire novel over-sampling techniques for learning from imbalanced datasets [22].

According to the availability of the output information (class label for classification and continuous value for regression), variable selection can be categorized as supervised variable selection [23,24], unsupervised variable selection [25,26] and semi-supervised variable selection [27,28]. Supervised variable selection determines relevant variables using the relationship with the corresponding output information on the training samples. Unsupervised variable selection evaluates variable relevance by exploiting data properties: variance, separability, and distribution. Semi-supervised variable selects a discriminative variable subset by utilizing both supervised and unsupervised training data. These variable selection algorithms can also be categorized into filters and wrappers [19,29,30]. Furthermore, the model structure can be taken into account [31,32]. The filter model relies on general characteristics of the training data to select some variables without involving any learning algorithm. The wrapper model requires one predetermined learning algorithm in variable selection and uses its performance to determine relevant variables. In our study, all the outputs are known. Thus, two different methods for supervised variable ranking and selection are implemented: the

Please cite this article as: H. Daassi-Gnaba, et al., Wood moisture content prediction using feature selection techniques and a kernel method, Neurocomputing (2016), http://dx.doi.org/10.1016/j.neucom.2016.09.005

Gram–Schmidt orthogonalization procedure [33–35] with a wrapper approach [19] and the Delta test method [23].

4.1. Gram–Schmidt orthogonalization procedure

This section is devoted to recall briefly the Gram–Schmidt (GS) orthogonalization procedure for ranking the variables of a model that is linear with respect to its parameters. In the context of machine learning, this method was first introduced in [33]. Afterwards, it has been widely implemented for various purposes. The GS procedure is an iterative method. For ranking a set of D candidate variables, it proceeds, at the first iteration, by estimating the relevance of each variable by computing the following quantities:

$$\cos^2(\mathbf{x}_k, \mathbf{y}) = \frac{\langle \mathbf{x}_k, \mathbf{y}\rangle^2}{\|\mathbf{x}_k\|^2 \|\mathbf{y}\|^2}, \quad k = 1, \ldots, D \tag{4}$$

where $\langle .,. \rangle$ denotes the dot product; $\mathbf{x}_k$ is the vector of the values of the k-th variable and $\mathbf{y}$ is the vector of output measured values of the process to be modeled.

The most relevant variable is the input vector that exhibits the largest value of this quantity. Projecting the $D-1$ remaining variables and the output vector $\mathbf{y}$ onto the subspace orthogonal to the vector of the most relevant variable ends the first iteration. Indeed, this projection permits to avoid the selection of redundant variables. The second iteration proceeds similarly by computing the relevance of the $D-1$ variables, selecting the most relevant. Projecting the $D-2$ remaining variables and the output vector $\mathbf{y}$ onto the subspace orthogonal to the first two ranked features. The procedure is repeated until all the variables are ranked. Once the features are ranked, the d most relevant of them can be selected using either a filter or a wrapper approach [19]. Although the filter approach is often computationally cost effective, the wrapper approach that usually leads to an acceptable computational burden in our implementation is preferred. Fig. 3 describes a basic example of the implementation of this procedure with two variables $\mathbf{x}_1$ and $\mathbf{x}_2$.

4.2. Delta test for variable selection

The Delta test method is a variable selection technique that is based on the estimation of noise variance [23]. Considering the set of N training samples $(\mathbf{x}_i, y_i)_{i=1}^N$, $\mathbf{x}_i \in \mathbb{R}^D$ and $y_i \in \mathbb{R}$, the relationship between the inputs and outputs is assumed to:

$$y_i = g(\mathbf{x}_i) + r_i \tag{5}$$

where g is the unknown function and $\mathbf{r} = (r_i)_{i=1}^N$ is the additive noise (with zero mean and finite variance). Noise variance estimation is the study of how to find an a priori estimate for $\mathrm{Var}(\mathbf{r})$ given some data without considering any specificities of the shape of g.

The Nearest Neighbour (NN) formulation of Delta test [36–38] estimates $\mathrm{Var}(\mathbf{r})$ calculated according to:

$$\mathrm{Var}(\mathbf{r}) \approx \delta = \frac{1}{2N}\sum_{i=1}^{N}\left(y_i - y_{\mathrm{NN}(i)}\right)^2 \quad \text{with} \quad \delta \to \mathrm{Var}(\mathbf{r}) \quad \text{as} \quad N \to \infty \tag{6}$$

where $y_{\mathrm{NN}(i)}$ is the output corresponding to input $\mathbf{x}_{\mathrm{NN}(i)}$ that is the NN of $\mathbf{x}_i \in \mathbb{R}^D$. $\mathbf{x}_{\mathrm{NN}(i)}$ is determined by minimizing the Euclidean distance: $\|\mathbf{x}_i - \mathbf{x}_j\|^2_{i\neq j}$ in the input space.

The Delta test is used as a cost function for variable selection. The selected variable subset is the one that best minimizes this score (6). In this study, a sequential forward selection is implemented; this method starts from an empty selection and proceeds by adding incrementally the variables that achieve the best improvement of the Delta test.

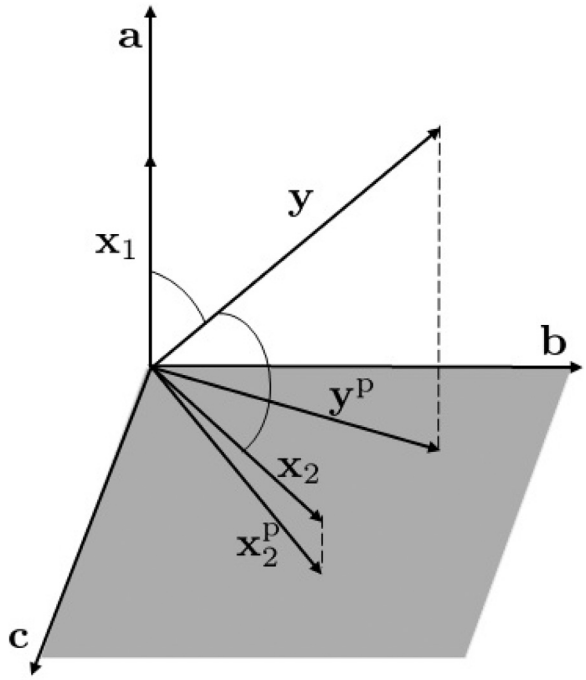


Fig. 3. The Gram–Schmidt orthogonalization procedure for ranking two candidate variables in a 3-dimensional vector space **abc**. At the first step, $\mathbf{x}_1$ is the most relevant variable with respect to the output $\mathbf{y}$. It is selected and then $\mathbf{y}$ and variable $\mathbf{x}_2$ are projected on the subspace orthogonal to $\mathbf{x}_1$. If there are more projected variables $\mathbf{x}_i^{\mathrm{p}}$, their relevance is computed with respect to $\mathbf{y}^{\mathrm{p}}$.

5. LS-SVM for linear and nonlinear regression

At the start of this section, basic information for LS-SVM theory is provided. Then, the strategy to optimize the hyperparameters of LS-SVM models is detailed.

5.1. LS-SVM model description

As in all machine learning problems, where a set of N training samples $(\mathbf{x}_i, y_i)_{i=1}^N$ is observed, $\mathbf{x}_i \in \mathbb{R}^D$ and $y_i \in \mathbb{R}$ denote input and output, respectively. The aim is to find a mapping function $h(\mathbf{x})$ so that $h(\mathbf{x}_i) \approx y_i \ \ \forall \ \ i$. In the case of SVM,

$$h(\mathbf{x}) = \underbrace{\langle \mathbf{w}, \varphi(\mathbf{x})\rangle + b}_{\hat{y}} \tag{7}$$

where $\varphi: \mathbb{R}^D \to \mathbb{R}^{D_k}$ is the nonlinear function that maps the input space to a high dimensional feature space where linear regression is possible; $\mathbf{w} \in \mathbb{R}^{D_k}$ is the weight vector; b is the bias term. The last two parameters are to be estimated from the training data. For the optimization problem, SVM involves inequality constraints and uses ϵ-insensitive loss function. The LS-SVM is a reformulation of the principles of SVM, that replaces inequality constraints by equality constraints. Furthermore, LS-SVM uses the Least Squares (LS) loss function. Therefore, in LS-SVM for model estimation, the optimization problem is given by

$$\begin{aligned} &\text{minimize} \quad J(\mathbf{w}, b, \mathbf{e}) = \frac{1}{2}\|\mathbf{w}\|^2 + C\sum_{i=1}^{N} e_i^2 \\ &y_i = \underbrace{\langle \mathbf{w}, \varphi(\mathbf{x}_i)\rangle + b}_{\hat{y}_i} + e_i, \quad i = 1, \ldots, N \end{aligned} \tag{8}$$

where e_i is the modeling error on example i and $C \geq 0$ is a regularization parameter. This problem can be solved using the Lagrange method:

Please cite this article as: H. Daassi-Gnaba, et al., Wood moisture content prediction using feature selection techniques and a kernel method, Neurocomputing (2016), http://dx.doi.org/10.1016/j.neucom.2016.09.005

$$\text{minimize} \quad L(\mathbf{w}, b, \mathbf{e}, \boldsymbol{\alpha}) = J(\mathbf{w}, b, \mathbf{e}) - \sum_{i=1}^{N} \alpha_i \left[\underbrace{\langle \mathbf{w}, \varphi(\mathbf{x}_i) \rangle + b}_{\hat{y}_i} + e_i - y_i \right] \tag{9}$$

with Lagrange multipliers $\alpha_i \in \mathbb{R}$. After elimination of $\mathbf{w}$ and $\mathbf{e}$, the application of Karush–Kuhn–Tucker conditions leads to linear system:

$$\begin{bmatrix} \mathbf{K} + \frac{1}{2C}\mathbf{I}_N & 1 \\ 1_N^T & 0 \end{bmatrix} \begin{bmatrix} \boldsymbol{\alpha} \\ b \end{bmatrix} = \begin{bmatrix} \mathbf{y} \\ 0 \end{bmatrix} \tag{10}$$

where $\mathbf{y} = [y_1; \ldots; y_N]$, $\boldsymbol{\alpha} = [\alpha_1; \ldots; \alpha_N]$, 1_N^T is the N-dimensional row vector whose elements are equal to 1, $\mathbf{I}_N$ is the $N \times N$ identity matrix and $\mathbf{K}$ is the $N \times N$ kernel matrix defined by:

$$k_{ij} = \langle \varphi(\mathbf{x}_i), \varphi(\mathbf{x}_j) \rangle \quad i, j = 1, \ldots, N$$

According to Mercer's condition, the inner product $\langle \varphi(\mathbf{x}_i), \varphi(\mathbf{x}_j) \rangle$ can be defined through a kernel function $K(\mathbf{x}_i, \mathbf{x}_j)$. Several choices for the kernel function are possible [39]. In this study, the following kernels were set:

- For linear regression, the linear kernel: $K(\mathbf{x}, \mathbf{x}_i) = \langle \mathbf{x}, \mathbf{x}_i \rangle$. The regularization parameter C is optimized using a validation procedure.
- For nonlinear regression, the Gaussian kernel: $K(\mathbf{x}, \mathbf{x}_i) = \exp\left(\frac{-\|\mathbf{x} - \mathbf{x}_i\|^2}{2\sigma^2} \right)$. The regularization parameter C and parameter σ are simultaneously optimized using a validation procedure. C and σ are called the hyperparameters of the optimization problem.

Hence, expression (7) becomes:

$$h(\mathbf{x}) = \sum_{i=1}^{N} \alpha_i K(\mathbf{x}, \mathbf{x}_i) + b \tag{11}$$

where the parameters of the LS-SVM model $\boldsymbol{\alpha}$ and b are the solution of the linear system given by relation (10).

5.2. LS-SVM model selection procedure

In the following, the strategy to optimize the hyperparameters of LS-SVM models presented above is described. A suitable way to proceed consists in selecting parameter C (for a linear regression) or the two parameters C and σ (for a nonlinear regression) that confer to the LS-SVM model the best generalization capabilities.

In practice, the hyperparameters are optimized by calculating a validation error. Several validation methods are mentioned in the literature [40]. The most popular procedures are the cross validation method and the Leave-One-Out (LOO) method. In order to reduce substantially the computational time of the selection procedure without compromising its efficiency, the validation error is estimated using the Virtual Leave-One-Out (VLOO) method. This method, first proposed for linear models [41] and later extended to nonlinear models [42], allows an estimation of the validation error to be computed by performing only one training involving the whole available data. This estimation is exact when dealing with linear-in-their-parameters models such as LS-SVM models. The implementation of the VLOO for LS-SVM models was recently described in [43]. For a given LS-SVM model, the VLOO error is computed as:

$$\text{VLOO} = \sqrt{\frac{1}{N} \sum_{i=1}^{N} \left(\frac{\alpha_i}{M_{ii}^{-1}} \right)^2} \tag{12}$$

where M_{ii}^{-1} is the i-th diagonal element of the inverse of matrix $\mathbf{M} = \mathbf{K} + \frac{1}{2C}\mathbf{I}_N$ that appears in the linear system of relation (10). Thus, the VLOO permits a fast and exact estimation of the validation error that consists in a great benefit when optimizing the values of the hyperparameters according to a grid search. This method was successfully applied to modeling with real data [44].

6. Neural networks

During the past years, the nonlinear modeling of processes using neural networks has been extensively studied [40,45]. The neural networks implemented in this study are feedforward nets with one hidden layer of sigmoidal units and direct connections from the inputs (see Fig. 4). The output of the network is given by

$$\Psi(\mathbf{x}, \boldsymbol{\theta}) = \sum_{i=1}^{N_c} \theta_i^{(2)} \tanh\left(\sum_{j=0}^{D} \theta_{ij}^{(1)} x_j \right) + \sum_{j=0}^{D} \theta_j^{(3)} x_j \tag{13}$$

where the input vector $\mathbf{x} \in \mathbb{R}^{D+1}$ is formed by $1, x_1, \ldots, x_D$. Parameter N_c is the number of hidden neurons and $\boldsymbol{\theta}$ is the vector of $N_p = (D + 1)(N_c + 1) + N_c$ parameters of the network. $\boldsymbol{\theta}$ is composed of $\boldsymbol{\theta}^{(1)}$, $\boldsymbol{\theta}^{(2)}$ and $\boldsymbol{\theta}^{(3)}$ components.

$\boldsymbol{\theta}$ is the set of parameters to be adjusted during the training phase. The Levenberg Marquardt algorithm [46,47] is used to minimize the traditional Least Squares (LS) cost function.

As in Section 5.2, the model selection procedure was also founded on the Virtual Leave-One-Out validation method (VLOO). For models that are nonlinear with respect to their parameters, the VLOO score [48,49] can be derived as:

$$\text{VLOO} = \sqrt{\frac{1}{N} \sum_{i=1}^{N} \left(\frac{e_i}{1 - h_{ii}} \right)^2} \tag{14}$$

where N is the total number of examples, $e_i = y_i - \Psi(\mathbf{x}_i, \boldsymbol{\theta})$ is the modeling error on example i and h_{ii} is the tangent plane leverage of example i. Tangent plane leverages for models that are nonlinear with respect to their parameters are an extension to those defined for models linear to their parameters [42,50]. They allow to estimate the influence of each sample of the training set on the parameters estimation. The tangent plane leverages are defined as

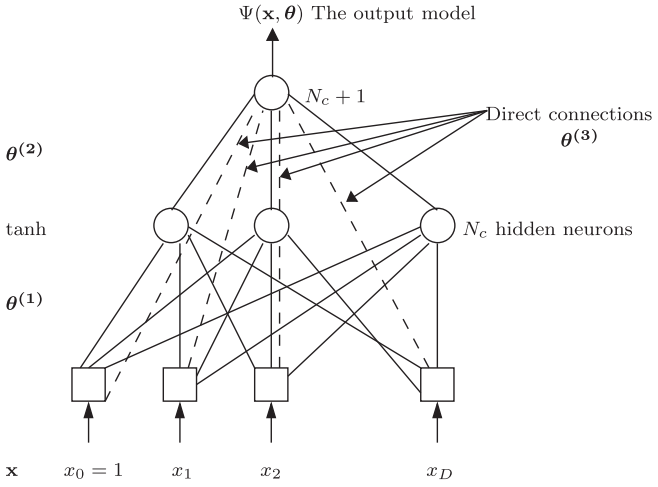


Fig. 4. A feedforward neural network with direct connections.

the diagonal elements of the orthogonal projection matrix **H** given by:

$$\mathbf{H} = \mathbf{Z}(\mathbf{Z}^T\mathbf{Z})^{-1}\mathbf{Z}^T \tag{15}$$

where **Z** is the Jacobian matrix of the model

$$Z_{ij} = \left(\frac{\partial \Psi(\mathbf{x}, \boldsymbol{\theta})}{\partial \theta_j}\right)_{\mathbf{x}=\mathbf{x}_i} \quad i = 1, \ldots, N \quad \text{and} \quad j = 1, \ldots, N_p \tag{16}$$

In practice, matrix **H** can be computed accurately using the Singular Value Decomposition (SVD) for matrix **Z** [51]. The tangent plane leverages come with remarkable properties: (i) their values lie in the interval [0, 1], (ii) their sum is equal to the rank of matrix **Z**.

Several neural nets were trained by increasing the number of hidden units beginning from 1. For each architecture, 100 training were performed with a different parameters initialization at each time. The model selection process consisted in determining the lowest number of hidden neurons N_c that best minimizes the VLOO score (14). Note that for linear models, the VLOO score can be computed similar to Eq. (14). However, the Jacobian matrix is replaced by the observation one.

The feedforward neural nets presented in this section are entirely deterministic. For various structured prediction problems, modeling techniques may be found on stochastic hidden variables instead of deterministic ones. In [52] a stochastic feedforward neural net called Sigmoid Belief net is proposed. This model is a kind of Bayesian net. It consists of a directed acyclic graph formed by nodes connected to each others by arcs. The nodes are random variables and the arcs represent probabilistic dependencies. To learn an optimal structure of Bayesian nets, original methods that improve the heuristic search algorithms were recently proposed in [53–55].

7. Design of the reflection coefficient and MC models

7.1. Building the reflection coefficient models

For each MC value, we have formed a set of M values for the frequency: $(f_i)_{i=1}^M$. For each value of the frequency, we have measured the corresponding reflection coefficient $(S_{11i})_{i=1}^M$. The experiment consisting in M measurements is repeated k times corresponding to k different wood chips arrangements. The whole measurements form a set of $N = M \times k$ samples that will be considered below as a training set: $(\mathbf{x}_i, y_i)_{i=1}^N = (f_i, S_{11i})_{i=1}^N$ where $f_i \in \mathbb{R}^D (D = 1)$ and $S_{11i} \in \mathbb{R}$. Measurements show that reflection coefficient S_{11} varies nonlinearly with respect to the frequency (see Fig. 1 (right)). This indicates that the LS-SVM technique is suitable to design the reflection coefficient models. While training and selecting these models, the values of the hyperparameters (C, σ) are simultaneously optimized according to a grid search (see Section 5.2). A set of candidate values for C and a range of σ are defined. For each value of (C, σ) and according to the training samples, the parameters $\boldsymbol{\alpha}$ of the model are calculated by solving the linear system (10), then the Virtual Leave-One-Out (VLOO) error is estimated using (12). The value of (C, σ) that best minimizes the VLOO error is selected. The Mean Square Error on the training samples (MSETrain) is calculated as follows:

$$\text{MSETrain} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(\underbrace{S_{11i} - \hat{S}_{11i}}_{e_i}\right)^2}. \tag{17}$$

The neural networks (see Section 6) are also implemented to design the reflection coefficient models and to compare results obtained with non linear LS-SVM.

7.2. Building the MC model

As described above, a reflection coefficient model was built for each value of MC and its predicted values $\hat{\mathbf{S}}_{\mathbf{11}}$ computed and saved. The later computation is performed for all the reflection coefficient models. The calculated values of $\hat{\mathbf{S}}_{\mathbf{11}}$ and the corresponding measurements MC lead to a set of N samples: $(\mathbf{x}_i, y_i)_{i=1}^N = ([\hat{\mathbf{S}}_{\mathbf{11}}]_i, \text{MC}_i)_{i=1}^N$ where $N = 16$, $[\hat{\mathbf{S}}_{\mathbf{11}}]_i \in \mathbb{R}^D (D = M = 801)$ and $\text{MC}_i \in \mathbb{R}$. This data will be used as a training set to build the MC model. This model also uses LS-SVM however with linear kernel since Gaussian kernel shows no improvement on the generalization capabilities. Contrarily to the set of models above, the number of input variables is large. Thus, the involvement of ranking and feature selection methods is desirable. The Gram–Schmidt orthogonalization procedure (see Section 4.1) is used to rank the $D = 801$ variables. Once the features are ranked, the d most relevant of them can be selected according to an incremental procedure. It starts by the most relevant variable $(d = 1)$, a set of candidate values for C is defined.

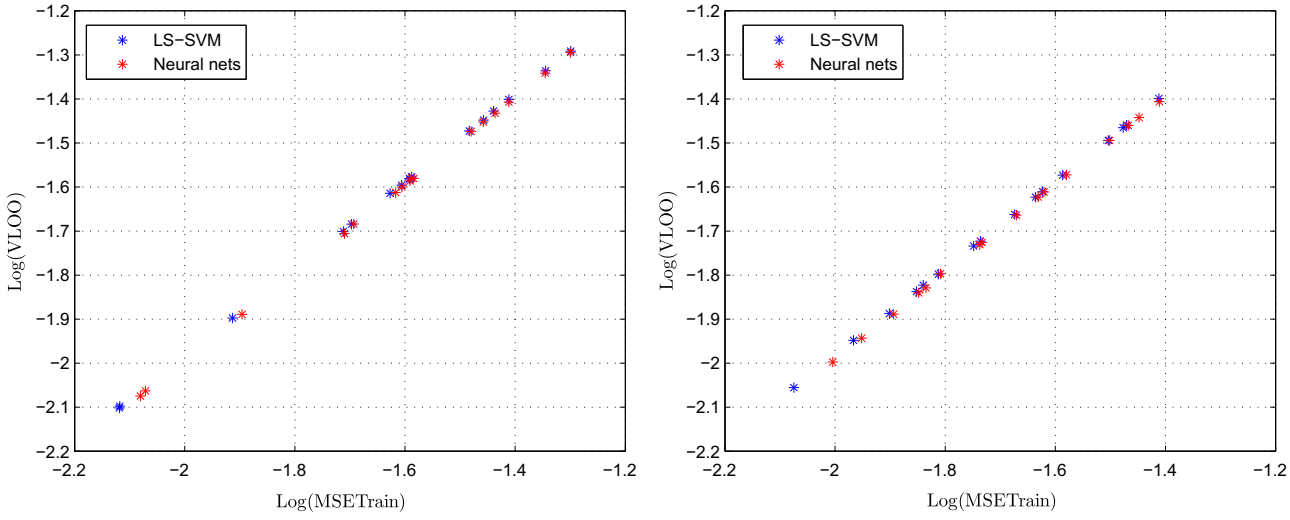


Fig. 5. Log(VLOO) values with respect to the Log(MSETrain) values for the reflection coefficient models when the half-wave dipole antenna 45 cm is applied for softwood (left) and hardwood (right). (For interpretation of the references to colour in this figure caption, the reader is referred to the web version of this paper.)

Please cite this article as: H. Daassi-Gnaba, et al., Wood moisture content prediction using feature selection techniques and a kernel method, Neurocomputing (2016), http://dx.doi.org/10.1016/j.neucom.2016.09.005

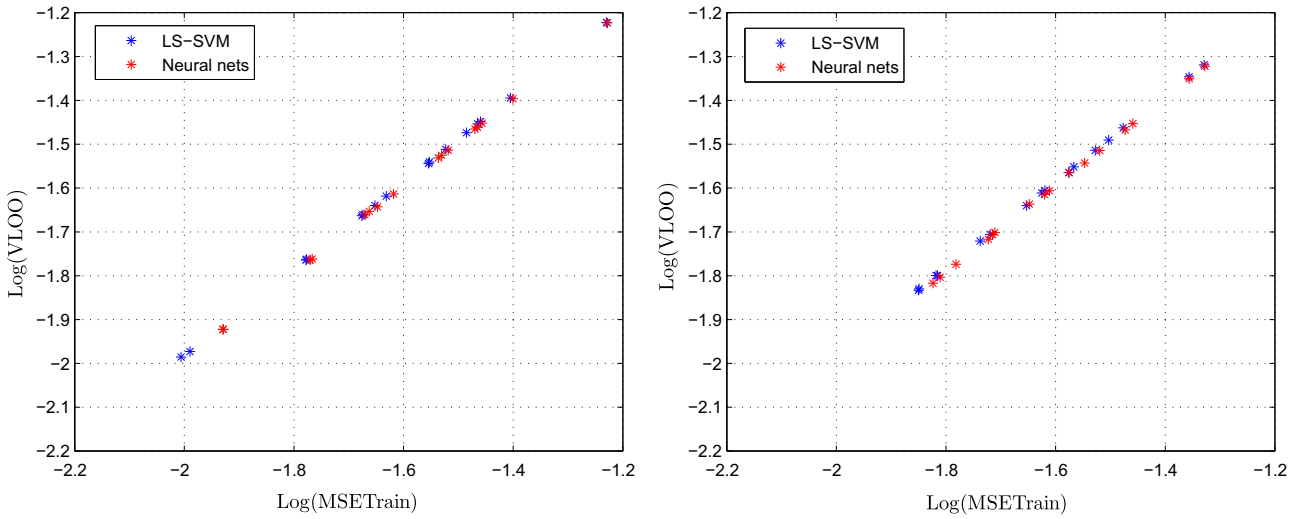


Fig. 6. Log(VLOO) values with respect to the Log(MSETrain) values for the reflection coefficient models when the half-wave dipole antenna 54 cm is applied for softwood (left) and hardwood (right).

Table 1
The MC models (dipole antenna 45 cm for softwood).

Methods	MSETrain	VLOO	C	d
LS-SVM with linear kernel	1.43	3.02	0.003	801
Gram–Schmidt + LS-SVM with linear kernel	**0.66**	**1.47**	**50**	**8**
Gram-Schmidt + LS	1.06	2.47	X	8
Delta-test + LS-SVM with linear kernel	2.22	2.93	3	3
Delta-test + LS	7.07	8.58	X	3

Table 2
The MC models (dipole antenna 45 cm for hardwood).

Methods	MSETrain	VLOO	C	d
LS-SVM with linear kernel	3.49	5.03	0.0008	801
Gram–Schmidt + LS-SVM with linear kernel	**1.52**	**2.60**	**2**	**7**
Gram–Schmidt + LS	**1.23**	**2.33**	**X**	**7**
Delta-test + LS-SVM with linear kernel	2.99	3.80	0.6	2
Delta-test + LS	4.55	5.30	X	2

Table 3
The MC models (dipole antenna 54 cm for softwood).

Methods	MSETrain	VLOO	C	d
LS-SVM with linear kernel	1.08	3.26	0.007	801
Gram–Schmidt + LS-SVM with linear kernel	**0.24**	**1.42**	**70**	**12**
Gram–Schmidt + LS	**0.21**	**1.20**	**X**	**12**
Delta-test + LS-SVM with linear kernel	2.51	3.23	0.2	5
Delta-test + LS	3.32	4.86	X	5

Table 4
The MC models (dipole antenna 54 cm for hardwood).

Methods	MSETrain	VLOO	C	d
LS-SVM with linear kernel	3.30	4.81	0.001	801
Gram–Schmidt + LS-SVM with linear kernel	**0.85**	**1.83**	**65**	**8**
Gram-Schmidt + LS	1.52	2.67	X	8
Delta-test + LS-SVM with linear kernel	3.93	4.92	0.1	5
Delta-test + LS	3.35	6.20	X	5

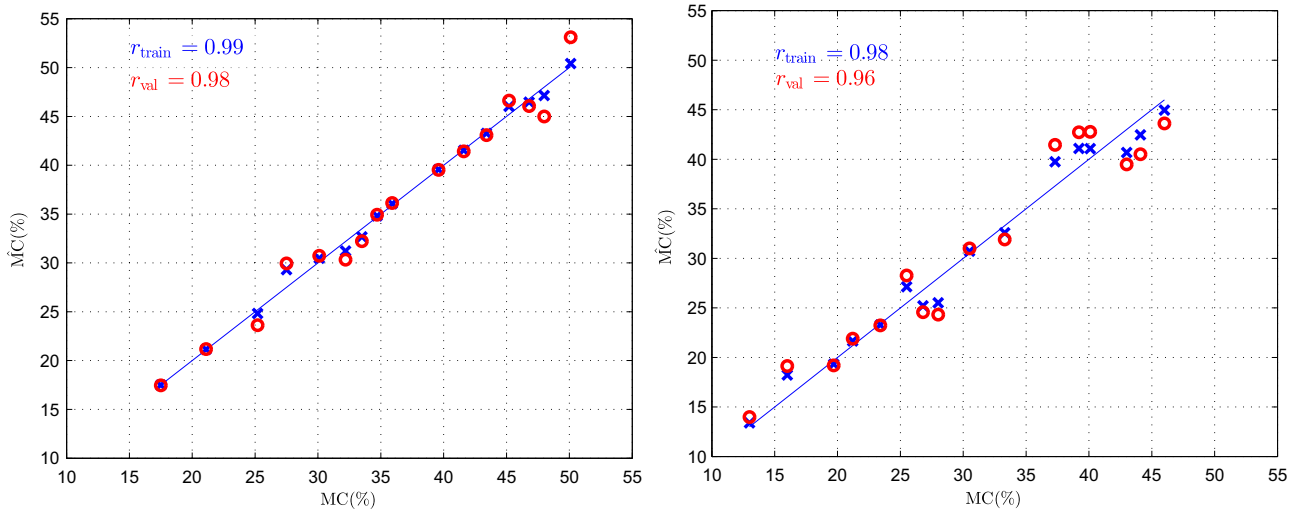


Fig. 7. The MC predicted values when Gram–Schmidt + LS-SVM with linear kernel is applied to softwood (left) and hardwood (right) with the half-wave dipole antenna 45 cm. (For interpretation of the references to colour in this figure caption, the reader is referred to the web version of this paper.)

Please cite this article as: H. Daassi-Gnaba, et al., Wood moisture content prediction using feature selection techniques and a kernel method, Neurocomputing (2016), http://dx.doi.org/10.1016/j.neucom.2016.09.005

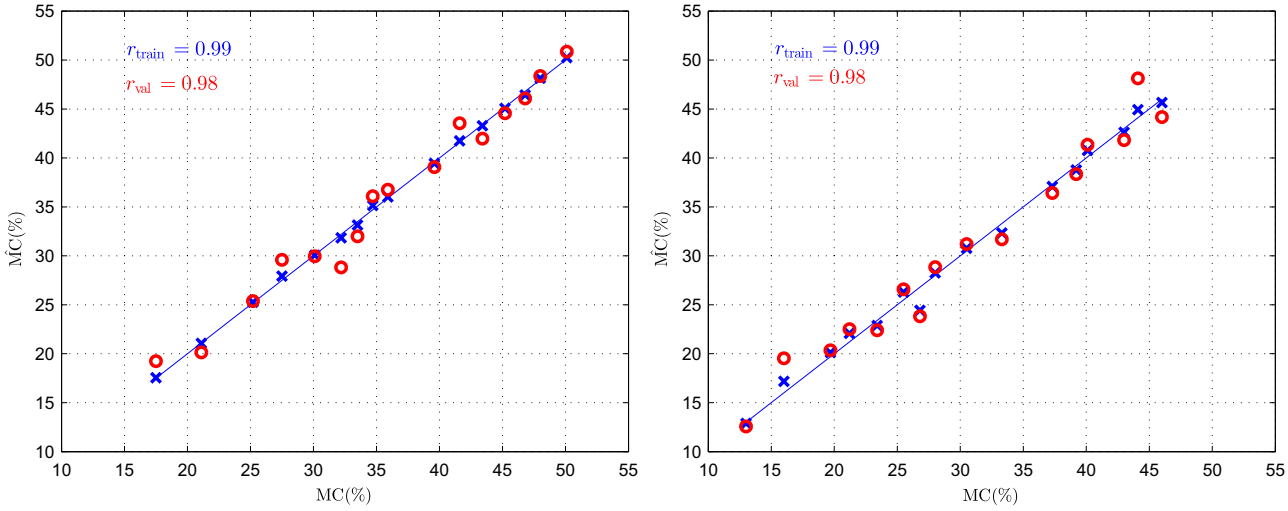


Fig. 8. The MC predicted values when Gram–Schmidt + LS-SVM with linear kernel is applied to softwood (left) and hardwood (right) with the half-wave dipole antenna 54 cm. (For interpretation of the references to colour in this figure caption, the reader is referred to the web version of this paper.)

Table 5
The MC models based on LS-SVM with linear kernel (dipole antenna 45 cm for mixture of wood).

Methods	MSETrain	VLOO	MSETest	C	d
LS-SVM with linear kernel	2.31	3.84	3.82	0.004	801
Gram–Schmidt + LS-SVM with linear kernel	**1.22**	**1.89**	**3.06**	**4**	**8**
Delta-test + LS-SVM with linear kernel	2.46	3.03	4.04	1000	4

Table 6
The MC models based on LS-SVM with linear kernel (dipole antenna 54 cm for mixture of wood).

Methods	MSETrain	VLOO	MSETest	C	d
LS-SVM with linear kernel	1.55	3.08	3.87	0.009	801
Gram–Schmidt + LS-SVM with linear kernel	**0.55**	**2.15**	**4.36**	**50**	**15**
Delta-test + LS-SVM with linear kernel	2.56	3.14	4.14	0.3	7

For each value of C and according to the training samples, the parameters $\boldsymbol{\alpha}$ of the model are calculated by solving the linear system (10), then the VLOO error is estimated using (12). The next most relevant variable ($d = 2$) is added. We repeat the same procedure while ranked variables are available. The value of (d,C) that best minimizes the VLOO error is selected. The MSETrain is calculated referring to relation (18):

$$\text{MSETrain} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(\underbrace{\text{MC}_i - \hat{\text{MC}}_i}_{e_i}\right)^2}. \tag{18}$$

A second feature selection method, the Delta test, described in Section 4.2 was also implemented to select the d most relevant variables according to the sequential forward selection; this method starts from an empty selection and by adding sequentially the variables that allow the best improvement of the Delta test (see Eq. (6)). When the subset of the d most relevant variables is determined, the training samples can be performed. A set of candidate values for C is defined. For each value of C and according to the training samples, the parameters $\boldsymbol{\alpha}$ of the model are calculated by solving the linear system (10),

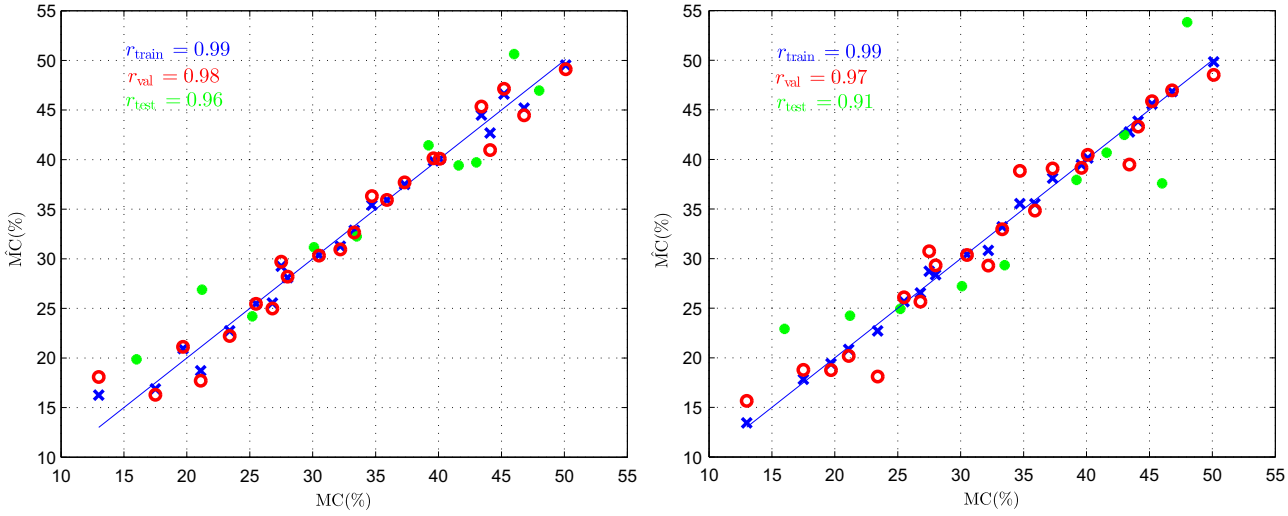


Fig. 9. The MC predicted values when Gram–Schmidt + LS-SVM with linear kernel is applied to mixture of wood with both antenna lengths 45 cm (left) and 54 cm (right). The training errors are in blue, validation errors are in red and test errors are in green. (For interpretation of the references to colour in this figure caption, the reader is referred to the web version of this paper.)

Please cite this article as: H. Daassi-Gnaba, et al., Wood moisture content prediction using feature selection techniques and a kernel method, Neurocomputing (2016), http://dx.doi.org/10.1016/j.neucom.2016.09.005

Table 7
The reflection coefficient models of the half-wave dipole antenna 45 cm for softwood (left) and hardwood (right) with LS-SVM.

MC (%)	MSETrain	VLOO	C	σ
17.5	0.0076	0.0080	10	0.03
21.1	0.0076	0.0079	90	0.04
25.2	0.0256	0.0263	3	0.05
27.5	0.0256	0.0262	50	0.07
30.1	0.0502	0.0511	10	0.08
32.2	0.0349	0.0356	30	0.07
33.5	0.0248	0.0254	20	0.06
34.7	0.0194	0.0199	5	0.05
35.9	0.0122	0.0127	20	0.04
39.6	0.0201	0.0207	30	0.05
41.6	0.0258	0.0265	5	0.05
43.4	0.0388	0.0397	20	0.06
45.2	0.0364	0.0374	5	0.05
46.8	0.0329	0.0337	20	0.06
48	0.0236	0.0243	40	0.05
50.1	0.0452	0.0461	1	0.06
13	0.0126	0.0130	80	0.05
16	0.0145	0.0150	60	0.04
19.7	0.0154	0.0159	10	0.04
21.2	0.0179	0.0185	5	0.04
23.4	0.0084	0.0088	10	0.03
25.5	0.0108	0.0113	5	0.03
26.8	0.0184	0.0189	60	0.05
28	0.0212	0.0218	5	0.05
30.5	0.0141	0.0145	10	0.04
33.3	0.0238	0.0245	20	0.05
37.3	0.0338	0.0347	5	0.05
39.2	0.0313	0.0320	5	0.06
40.1	0.0231	0.0238	30	0.05
43	0.0259	0.0267	90	0.05
44.1	0.0387	0.0399	40	0.05
46	0.0334	0.0342	5	0.05

Table 8
The reflection coefficient models of the half-wave dipole antenna 45 cm for softwood (left) and hardwood (right) with neural network.

MC (%)	MSETrain	VLOO	N_c
17.5	0.0085	0.0086	10
21.1	0.0083	0.0084	10
25.2	0.0257	0.0259	10
27.5	0.0258	0.0263	10
30.1	0.0502	0.0508	9
32.2	0.0349	0.0353	10
33.5	0.0248	0.0251	10
34.7	0.0195	0.0197	9
35.9	0.0127	0.0129	10
39.6	0.0203	0.0207	10
41.6	0.0260	0.0263	9
43.4	0.0388	0.0391	9
45.2	0.0366	0.0370	10
46.8	0.0331	0.0336	9
48	0.0242	0.0244	10
50.1	0.0451	0.0456	10
13	0.0128	0.0129	10
16	0.0146	0.0148	10
19.7	0.0155	0.0160	10
21.2	0.0183	0.0186	10
23.4	0.0099	0.0101	10
25.5	0.0112	0.0114	10
26.8	0.0185	0.0188	10
28	0.0214	0.0217	10
30.5	0.0142	0.0144	10
33.3	0.0240	0.0245	10
37.3	0.0341	0.0347	10
39.2	0.0315	0.0320	9
40.1	0.0234	0.0238	10
43	0.0263	0.0268	10
44.1	0.0388	0.0393	10
46	0.0356	0.0361	10

then the VLOO error is estimated using (12). The value of C that best minimizes the VLOO error is selected.

With both methods, linear model based on traditional Least Squares (LS) is also used to build the MC model. The obtained results are compared with those achieved when the linear LS-SVM is applied.

8. Experimental results

8.1. Reflection coefficient models

Two wood species were involved: softwood (pine wood) and hardwood (oak wood). One pile of each species was available, it was first moistened and thereafter the change in the MC was monitored over time. 16 different MC values were tested. For each MC value, $M=801$ values of the reflection coefficient S_{11} were measured corresponding to as many different frequencies regularly spaced in the domain [50 MHz, 1.3 GHz]. $k=4$ different measurements of the reflection coefficient S_{11} corresponding to 4 different wood chips arrangements were performed. In addition, we studied and compared the performance of dipole having 45 cm and 54 cm lengths, respectively.

8.1.1. *Half-wave dipole antenna* 45 cm

The reflection coefficient S_{11} of the half-wave dipole antenna 45 cm was measured at 16 different MC values for softwood and hardwood. Therefore, for each MC value, a reflection coefficient model was designed. Thus, 16 reflection coefficient models for softwood and 16 reflection coefficient models for hardwood were established. Fig. 5 illustrates the Log(VLOO) values with respect to the Log(MSETrain) values for these models obtained with both the LS-SVM technique (blue points) and neural nets (red points). This figure shows that for each model, the training and validation errors (MSETrain, VLOO) are very close for both methods. This result ensures that the models complexity is well controlled. Thus, no overfitting is observed. LS-SVM models show slightly better results probably due to the built-in regularization mechanism. Further details can be found in the appendix:

- Table 7 gives the training and validation errors (MSETrain, VLOO) and also the couple (C, σ) of the LS-SVM models showing the best generalization capabilities for each value of MC and for both wood types.
- Table 8 gives the (MSETrain, VLOO) and the optimal hidden neurons N_c of the neural networks models for both wood types.

8.1.2. *Half-wave dipole antenna* 54 cm

When the half-wave dipole antenna 54 cm was used, the same process to determine the reflection coefficient models was carried out. For each wood type, 16 reflection coefficient models were selected. Fig. 6 illustrates the Log(VLOO) values with respect to the Log(MSETrain) values for these models. Similarly to the antenna 45 cm, this figure shows that for each model, the training and validation errors (MSETrain, VLOO) are very close with both modeling methods. No overfitting is observed. LS-SVM also perform slightly better than above probably thanks to a good control of models complexity with the built-in regularization mechanism. Further details can be found in the appendix:

- Table 9 illustrates the training and validation errors (MSETrain, VLOO) and also the couple (C, σ) of the models showing the best generalization capabilities for each value of MC and for both wood types.
- Table 10 gives the (MSETrain, VLOO) and the optimal hidden neurons N_c of the neural networks models for both wood types.

Please cite this article as: H. Daassi-Gnaba, et al., Wood moisture content prediction using feature selection techniques and a kernel method, Neurocomputing (2016), http://dx.doi.org/10.1016/j.neucom.2016.09.005

Table 9
The reflection coefficient models of the half-wave dipole antenna 54 cm for softwood (left) and hardwood (right) with LS-SVM.

MC (%)	MSETrain	VLOO	C	σ
17.5	0.0102	0.0107	60	0.04
21.1	0.0099	0.0103	20	0.03
25.2	0.0347	0.0356	5	0.05
27.5	0.0211	0.0217	40	0.05
30.1	0.0167	0.0172	110	0.05
32.2	0.0167	0.0173	5	0.04
33.5	0.0343	0.0352	4	0.05
34.7	0.0223	0.0229	10	0.05
35.9	0.0234	0.0241	30	0.05
39.6	0.0212	0.0218	60	0.05
41.6	0.0279	0.0286	30	0.06
43.4	0.0394	0.0403	20	0.06
45.2	0.0280	0.0288	10	0.05
46.8	0.0589	0.0599	70	0.09
48	0.0300	0.0308	5	0.05
50.1	0.0327	0.0336	5	0.05
13	0.0153	0.0159	20	0.04
16	0.0237	0.0245	5	0.04
19.7	0.0191	0.0197	4	0.04
21.2	0.0240	0.0248	6	0.04
23.4	0.0142	0.0148	10	0.03
25.5	0.0141	0.0147	70	0.04
26.8	0.0222	0.0229	30	0.05
28	0.0183	0.0190	90	0.04
30.5	0.0265	0.0273	7	0.05
33.3	0.0440	0.0451	1	0.05
37.3	0.0152	0.0159	5	0.03
39.2	0.0271	0.0281	20	0.04
40.1	0.0334	0.0345	5	0.04
43	0.0297	0.0306	4	0.04
44.1	0.0469	0.0480	10	0.06
46	0.0314	0.0323	30	0.05

Table 10
The reflection coefficient models of the half-wave dipole antenna 54 cm for softwood (left) and hardwood (right) with neural network.

MC (%)	MSETrain	VLOO	N_c
17.5	0.0118	0.0120	10
21.1	0.0118	0.0120	10
25.2	0.0349	0.0353	10
27.5	0.0214	0.0218	10
30.1	0.0170	0.0172	10
32.2	0.0171	0.0173	10
33.5	0.0343	0.0346	10
34.7	0.0225	0.0228	10
35.9	0.0241	0.0244	9
39.6	0.0217	0.0222	10
41.6	0.0291	0.0294	10
43.4	0.0398	0.0402	10
45.2	0.0295	0.0298	10
46.8	0.0592	0.0598	9
48	0.0303	0.0306	10
50.1	0.0339	0.0342	10
13	0.0165	0.0168	10
16	0.0240	0.0243	10
19.7	0.0195	0.0199	10
21.2	0.0245	0.0248	10
23.4	0.0155	0.0157	10
25.5	0.0150	0.0152	10
26.8	0.0225	0.0231	10
28	0.0193	0.0197	10
30.5	0.0266	0.0272	10
33.3	0.0440	0.0446	10
37.3	0.0189	0.0192	10
39.2	0.0284	0.0287	9
40.1	0.0347	0.0352	10
43	0.0302	0.0306	10
44.1	0.0470	0.0476	9
46	0.0336	0.0341	9

8.2. MC *model*

The conditions of modeling stay the same as above. Two dipole lengths: 45 cm and 54 cm, two wood species: softwood and hardwood, and 16 different MC values for each wood species were considered. Depending on whether: (i) a feature selection procedure is implemented or not for building a MC model, (ii) either the Gram–Schmidt procedure or the Delta test is used, three different training procedures can be carried out. These procedures lead to three different MC models that are based on LS-SVM technique. All of them were taken into account.

8.2.1. *Half-wave dipole antenna* 45 cm

Tables 1 and 2 show the training and validation errors (MSETrain, VLOO) obtained with the three kinds of models. The selected models are those with the best generalization capabilities. For each of them, the tables also indicate the corresponding value of the regularization parameter C and the number d of selected relevant variables. Comparisons with the traditional LS method are also given.

For the softwood, the best performance was obtained when using the Gram–Schmidt procedure and LS-SVM technique with linear kernel. When the MC range is from 17.5% to 50.1%, the validation error is 1.47%. For the hardwood, the best performance was obtained with the same combination as for the softwood. When the MC range is from 13% to 46%, the validation error is 2.60% (almost 2.33% with LS). Fig. 7 illustrates the best results for both wood types. Training (blue) and validation (red) results are presented by scatter plots of the 16 predicted $\hat{\mathrm{MC}}$ values versus the 16 measured MC values. The predicted values are shown on the y-axis and the measured values on the x-axis. The quality of the model is also evaluated using the correlation coefficient r. This coefficient is computed as follows:

$$r = \frac{\sum_{i=1}^{N}\left(\mathrm{MC}_i - \overline{\mathrm{MC}}\right)\left(\hat{\mathrm{MC}}_i - \overline{\hat{\mathrm{MC}}}\right)}{\sqrt{\sum_{i=1}^{N}\left(\mathrm{MC}_i - \overline{\mathrm{MC}}\right)^2}\sqrt{\sum_{i=1}^{N}\left(\hat{\mathrm{MC}}_i - \overline{\hat{\mathrm{MC}}}\right)^2}} \tag{19}$$

where $\overline{\mathrm{MC}}$ and $\overline{\hat{\mathrm{MC}}}$ are the mean values of $(\mathrm{MC}_i)_{i=1}^{N}$ and $(\hat{\mathrm{MC}})_{i=1}^{N}$, respectively. The best quality of the model is evaluated with the highest value of r.

8.2.2. *Half-wave dipole antenna* 54 cm

The results obtained are shown in Tables 3 and 4 for softwood and hardwood, respectively.

Similarly to the antenna 45 cm, the best results are obtained when using the Gram–Schmidt procedure and LS-SVM technique with linear kernel. The validation error for the softwood is 1.42% (almost 1.20% with LS) and for the hardwood is 1.83%. Fig. 8 illustrates the best results for both wood types. Training (blue) and validation (red) results are presented by scatter plots of the 16 predicted $\hat{\mathrm{MC}}$ values with respect to the measured MC values.

8.2.3. *Discussion*

- Over the four combinations obtained with the two dipole lengths and the two wood types, the most efficient model is achieved when the reflection coefficients are selected using the Gram–Schmidt orthogonalization procedure.
- The half-wave dipole antenna 54 cm allows to achieve more efficient models than the dipole 45 cm whatever the wood type.
- The LS-SVM with Gaussian kernel was also implemented. The

Please cite this article as: H. Daassi-Gnaba, et al., Wood moisture content prediction using feature selection techniques and a kernel method, Neurocomputing (2016), http://dx.doi.org/10.1016/j.neucom.2016.09.005

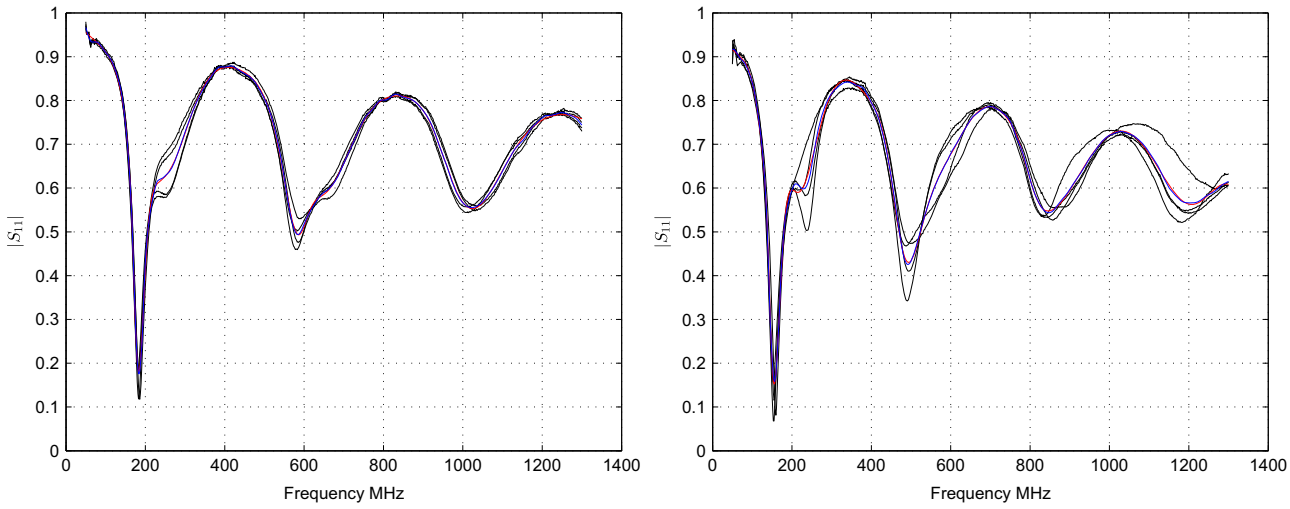


Fig. 10. The measurements of the reflection coefficient S_{11} (black), predicted values with LS-SVM (blue) and with neural network (red) for MC = 30% (left) and MC = 48% (right) with the 54 cm length antenna for softwood. (For interpretation of the references to colour in this figure caption, the reader is referred to the web version of this paper.)

values of the hyperparameters (C, σ) were simultaneously optimized using the VLOO procedure. The performance was not improved. Therefore, LS-SVM with linear kernel remains more suitable.

- Both LS-SVM and the traditional LS method perform pretty much the same when variables are ranked using the Gram–Schmidt orthogonalizatoin procedure. However, LS-SVM models outperform systematically the LS method when using the Delta-test.

8.2.4. Mixture of wood

The proposed measurement system shows good performances to estimate the MC of wood chips being either softwood or hardwood. However, a woodpile may consist in a mixture of these two wood types. Thus, one can wonder if the modeling methodology proposed may achieve efficient models when the design uses measurements obtained with both wood types. In Section 8.1, we built 16 reflection coefficient models for softwood and 16 reflection coefficient models for hardwood for both antenna lengths. This leads to 32 reflection coefficient models corresponding to 32 different values of MC. Consequently, 32 samples $(|\hat{\mathbf{S}}_{\mathbf{11}}|_i, \mathrm{MC}_i)_{i=1}^{32}$ were formed. 22 samples (11 for softwood and 11 for hardwood) were dedicated to adjust the MC model. The remaining 10 samples were used to test the model. Depending on whether: (i) a feature selection procedure is implemented or not for building a MC model, (ii) either the Gram–Schmidt procedure or the Delta test is used, three different training procedures can be carried out. These procedures lead to three different MC models that are based on LS-SVM technique. All of them were taken into account. Tables 5 and 6 show the results obtained with the three kinds of models for dipole lengths 45 cm and 54 cm, respectively.

Similarly to the design of MC models with data coming from either softwood or hardwood, the best results when the two wood types are mixed were obtained using the Gram–Schmidt procedure and the LS-SVM technique with linear kernel. It is worth noting the improvement brought by variable selection methods. With dipole antenna 45 cm a LS-SVM model using small subset formed by $(d=8)$ most relevant variables allows to achieve when the MC range is from 13% to 50.1% a validation error of 1.89%, two times lower than a model using all the available variables. Fig. 9 illustrates the scatter plots of training data (blue), validation data (red) and test data (green) for the best models obtained with both antenna lengths.

9. Conclusion

To contribute to the wood-to-energy industrial progress a combined model for woody biomass Moisture Content (MC) prediction was proposed. The model design is based on RF measurements obtained with an antenna that is fully buried into the samples of wood chips. In order to take into account the variations caused by the heterogeneity of the samples, several measurements of the reflection coefficient S_{11} in the frequency-domain were carried out for the same sample. These measurements differ by the arrangement of the wood chips. The proposed modeling approach is implemented according to three steps: (i) perform the measurements, (ii) using all the available RF measurements, build reflection coefficient models to estimate S_{11} as a function of the frequency, (iii) use the outputs of the reflection coefficient models as inputs of the MC model. Thereby, the MC model allows to predict MC from estimate values of S_{11}. The reflection coefficient models are based on the LS-SVM technique with Gaussian kernel. The MC model also uses LS-SVM however with linear kernel since nonlinear kernel shows no improvement on the generalization capabilities. Variable ranking and selection methods were implemented and their performances compared. Using real world data obtained with two species of wood chips and two different antenna lengths, several numerical experiments were conducted. Results show the effectiveness of the proposed modeling methodology and the limited computational power required for its implementation. This novel solution for MC prediction is suitable for direct implementation on real-time wood-to-energy industrial processes. Indeed, an industrial prototype is currently being tested in a production site.

Acknowledgement

This work was supported by the French National Research Agency (ANR) as part of the "Bio-Matières et Energies (Bio-ME) 2012" Programme (ANR-12-BIME-0007).

Appendix

Details that may be useful to readers are provided in this section.

- Tables 7 and 9 illustrate the training and validation errors (MSETrain, VLOO) and also the couple (C, σ) of the models showing the best generalization capabilities for each value of MC and for both wood types when using two dipole lengths 45 cm and 54 cm, respectively.
- Tables 8 and 10 give the training and validation errors

(MSETrain, VLOO) and the optimal hidden neurons N_c of the neural networks models for both wood types when using two dipole lengths 45 cm and 54 cm, respectively.

Fig. 10 shows the measurements of the reflection coefficient S_{11} and the predicted values obtained with both modeling methods for two MC values.

References

[1] Analysis of options to move beyond 20% greenhouse gas emission reductions and assessing the risk of carbon leakage, Communication from the commission to the European Parliament, the council, the European economic and social Committee and the committee of regions, Brussels, 2010.
[2] F. Manzano-Agugliaro, A. Alcayde, F.G. Montoya, A. Zapata-Sierra, C. Gil, Scientific production of renewable energies worldwide: an overview, Renew. Sustain. Energy Rev. 18 (1) (2013) 134–143.
[3] P. Jensen, H. Hartmann, T. Böhm, M. Temmerman, F. Rabier, M. Morsing, Moisture content determination in solid biofuels by dielectric and NIR reflection methods, Biomass Bioenergy 30 (11) (2006) 935–943.
[4] H. Hartmann, T. Böhm, Rapid moisture content determination of wood chips – results from comparative trials, in: Proceedings of the 1st World Conference on Biomass for Energy and Industry, Sevilla, James & James Ltd., London, 2001, pp. 571–574.
[5] European Committee for Standardization, CENT/TC 335 EN 14774-1:2009 Determination of moisture content.
[6] R. Samuelsson, J. Burvall, R. Jirjis, Comparison of different methods for the determination of moisture content in biomass, Biomass Bioenergy 30 (2006) 929–934.
[7] C. Jun, Z. Jiawei, S. Liping, Comparison on prediction wood moisture content using ARIMA and improved neural networks, International Conference on Computational Intelligence for Measurement Systems and Applications (CIMSA), IEEE, Hong Kong, China, 2009, pp. 148–152.
[8] D. Zhang, J. Cao, L. Sun, Soft sensor modeling of moisture content in drying process based on LSSVM, The 9th International Conference on Electronic Measurement and Instruments (ICEMI), IEEE, Beijing, China, 2009, pp. 2-989–2-993.
[9] S. Wen, M. Deng, A. Inoue, Moisture content prediction of wood drying process using svm-based model, Int. J. Innov. Comput. Inf. Control (ICIC Int.) 8 (6) (2012) 4083–4093.
[10] J. Nyström, E. Dahlquist, Methods for determination of moisture content in woodchips for power plants a review, Fuel 83 (7–8) (2004) 773–779.
[11] J. Nyström, E. Thorin, S.O. Backa, E. Dahlquist, Moisture content measurements on sawdust with radio frequency spectroscopy, in: American Society of Mechanical Engineers (ASME) Power Conference, Chicago, USA, 2005, pp. 697–702.
[12] S. Avramidis, L. Iliadis, Prediction wood thermal conductivity using artificial neural networks, Wood Fiber Sci. 37 (4) (2005) 682–690.
[13] L. Iliadis, S. Tachos, S. Avramidis, S.D. Mansfield, Hybrid e-regression and validation soft computing techniques: the case of wood dielectric loss factor, Neurocomputing 107 (2013) 33–39.
[14] S. Ramo, J.R. Whinnery, T. Van. Duzer, Fields and Waves in Communication Electronics, John Wiley & Sons, New York, 1984.
[15] U. Kaatze, V. Uhlendorf, The dielectric properties of water at microwave frequencies, Z. Phys. Chem. N. F. 126 (5) (1981) 151–165.
[16] P. Mabire, A. Mayer, S. Holé, T. Ditchi, E. Géron, Z. Mokthari, Mesure non-destructive du taux d'humiditéd'un matériau, Patent FR1552032, 2015.
[17] M. Merlan, T. Ditchi, S. Holé, Y. Oussar, J. Lucas, E. Géron, Système électromagnétique de mesure de l'humiditéde plaquettes forestières, Journées Scientifiques de l'Union Radio-scientifique Internationale, Paris, France, 2015.
[18] M. Merlan, T. Ditchi, J. Lucas, E. Géron, Y. Oussar, S. Holé, Electromagnetic system to determine the moisture content of wood chips, in: Third International Congress on Energy Efficiency and Energy Related Materials (ENEFM), Oludeniz, Fethiye/Mugla, Turkey, 2015.
[19] I. Guyon, A. Elisseeff, An Introduction to variable and feature selection, J. Mach. Learn. Res. 3 (2003) 1157–1182.
[20] Y. Oussar, I. Ahriz, B. Denby, G. Dreyfus, Indoor localization based on cellular telephony RSSI fingerprints containing very large numbers of carriers, EURASIP J. Wirel. Commun. Netw. 2011 (1) (2011) 1–14.
[21] H. Daassi-Gnaba, Y. Oussar, External vs. internal SVM-RFE: the SVM-RFE method revisited and applied to emotion recognition, Neural Netw. World 25 (1) (2015) 75–91.
[22] X. Fan, K. Tang, T. Weise, Margin-based over-sampling method for learning from imbalanced datasets, in: Proceedings of the 15th Pacific-Asia Conference on Knowledge Discovery and Data Mining (PAKDD), Shenzhen (China), Part II, Lecture Notes in Computer Science, Lecture Notes in Artificial Intelligence, vol. 6635, 2011, pp. 309–320.
[23] E. Eirola, E. Liitiäinen, A. Lendasse, F. Corona, M. Verleysen, Using the delta test for variable selection, in: European Symposium on Artificial Neural Networks (ESANN), Bruges, Belgium, 2008, pp. 25–30.
[24] F. Nie, H. Huang, X. Cai, C. Ding, Efficient and robust feature selection via joint l2,1-norms minimization, in: Advances in Neural Information Processing Systems (NIPS), MIT Press, Cambridge, MA, USA, 2013, pp. 1813–1821.
[25] J.G. Dy, C.E. Brodley, Feature selection for unsupervised learning, J. Mach. Learn. Res. 5 (2004) 845–889.
[26] F. Nie, S. Xiang, Y. Jia, C. Zhang, S. Yan, Trace ratio criterion for feature selection, in: Proceedings of the 23th AAAI Conference on Artificial Intelligence, Chicago, USA, 2008, pp. 671–676.
[27] Y. Liu, F. Nie, J. Wu, L. Chen, Efficient semi-supervised feature selection with noise insensitive trace ratio criterion, Neurocomputing 105 (2013) 12–18.
[28] J. Zhao, K. Lu, X. He, Locality sensitive semi-supervised feature selection, Neurocomputing 71 (2008) 1842–1849.
[29] S. Das, Filters, wrappers and a boosting-based hybrid for feature selection, in: Proceedings of the 18th International Conference on Machine Learning (ICML), Williamstown, USA, 2001, pp. 74–81.
[30] Md.M. Kabir, Md.M. Islam, K. Murase, A new wrapper feature selection approach using neural network, Neurocomputing 73 (2010) 3273–3283.
[31] J. Weston, S. Mukherjee, O. Chapelle, M. Pontil, T. Poggio, V. Vapnik, Feature selection for svms, in: Neural Information Processing Systems, MIT Press, Cambridge, MA, 2001.
[32] J.R. Quevedo, A. Bahamonde, O. Luaces, A simple and efficient method for variable ranking according to their usefulness for learning, Comput. Stat. Data Anal. 52 (1) (2007) 578–595.
[33] S. Chen, S.A. Billings, W. Luo, Orthogonal least squares methods and their application to non-linear system identification, Int. J. Control 50 (5) (1989) 1873–1896.
[34] Y. Oussar, G. Dreyfus, Initialization by selection for wavelet network training, Neurocomputing 34 (2000) 131–143.
[35] H. Stoppiglia, G. Dreyfus, R. Dubois, Y. Oussar, Ranking a random feature for variable and feature selection, J. Mach. Learn. Res. 3 (2003) 1399–1414.
[36] H. Pi, C. Peterson, Finding the embedding dimension and variable dependencies in time series, Neural Comput. 6 (1994) 509–520.
[37] Aalbjörn Stefânson, N. Koncar, A.J. Jones, A note on the gamma test, Neural Comput. Appl. 5 (3) (1997) 131–133.
[38] E. Liitiäinen, M. Verleysen, F. Corona, A. Lendasse, Residual variance estimation in machine learning, Neurocomputing 72 (16–18) (2009) 3692–3703.
[39] C. Zhang, F. Nie, S. Xiang, A general kernelization framework for learning algorithms based on kernel PCA, Neurocomputing 73 (4–6) (2010) 959–967.
[40] T. Hastie, R. Tibshirani, J. Friedman, The Elements of Statistical Learning, Springer, New York, 2009.
[41] D.A. Belsley, E. Kuh, R.E. Welsh, Regression diagnostics: identifying influential data and sources of collinearity, John Wiley & Sons, New York, 1980.
[42] R. Laurent, R. Cook, Leverage, local influence and curvature in nonlinear regression, Biometrika 80 (1) (1993) 99–106.
[43] G.C. Cawley, N.L.C. Talbot, Preventing over-fitting during model selection via Bayesian regularization of the hyperparameters, J. Mach. Learn. Res. 8 (2007) 841–861.
[44] R. Sivira, H. Brogniez, C. Mallet, Y. Oussar, A layer-averaged relative humidity profile retrieval for microwave observations: design and results for the Megha–Tropiques payload, Atmos. Meas. Tech. 8 (3) (2015) 1055–1071.
[45] G. Dreyfus, Neural Networks Methodology and Applications, Springer-Verlag, Berlin, 2004.
[46] K. Levenberg, A method for the solution of certain non-linear problems in least squares, Q. J. Appl. Math. II 2 (1944) 164–168.
[47] D.W. Marquardt, An algorithm for least-squares estimation of nonlinear parameters, J. Soc. Ind. Appl. Math. 11 (2) (1963) 431–441.
[48] G. Monari, G. Dreyfus, Withdrawing an example from the training set: an analytic estimation of its effect on a non-linear parameterised model, Neurocomputing 35 (2000) 195–201.
[49] G. Monari, G. Dreyfus, Local overfitting control via leverages, Neural Comput. 14 (2002) 1481–1506.
[50] S. Chatterjee, A. Hadi, Influential observations, high leverage points, and outliers in linear regression, Stat. Sci. 1 (3) (1986) 379–393.
[51] G. Golub, C. Van Loan, Matrix Computations, The Johns Hopkins University Press, Maryland, 1996.
[52] R.M. Neal, Connectionist learning of belief networks, Artif. Intell. 56 (1992) 71–113.
[53] X. Fan, B. Malone, C. Yuan, Finding optimal Bayesian network structures with constraints learned from data, in: Proceedings of the 30th Conference on Uncertainty in Artificial Intelligence (UAI), Quebec, Canada, 2014, pp. 200–209.
[54] X. Fan, C. Yuan, B. Malone, Tightening bounds for Bayesian network structure learning, in Proceedings of the 28th AAAI Conference on Artificial Intelligence, Quebec, Canada, 2014, pp. 2439–2445.
[55] X. Fan, C. Yuan, An improved lower bound for Bayesian network structure learning, in: Proceedings of the 29th AAAI Conference on Artificial Intelligence, Austin, USA, 2015, pp. 3526–3532.

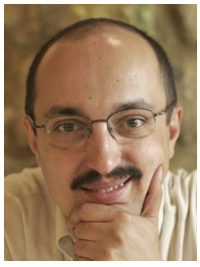
Hela Daassi was born in Sousse, Tunisia, in 1976. She graduated from the engineering school Ecole Nationale d'Ingénieurs de Tunis in 1999 and obtained his Ph.D. in Computer Science in2006 from the University of Paris 5, France. She is Postdoctoral researcher at ESPCI since 2015. She has made many contributions to speech processing in mobile networks and Machine Learning applied to emotion recognition from speech. Currently she works for machine learning techniques and their applications in various domains of instrumentation. ESPCI/UPMC/CNRS-LPEM/UMR8213, 10 rue Vauquelin, 75005 Paris, France. (email: hela.daassi@espci.fr).

Please cite this article as: H. Daassi-Gnaba, et al., Wood moisture content prediction using feature selection techniques and a kernel method, Neurocomputing (2016), http://dx.doi.org/10.1016/j.neucom.2016.09.005

Thierry Ditchi was born in Noisy le Sec, France, in 1959. He received his Ph.D. in electrical engineering in 1990 from the University Pierre et Marie Curie in association with the Ecole Supérieure de Physique et de Chimie Industrielles, Paris France. During his Ph.D. thesis, he studied electrical properties of insulating materials and developed instruments for measuring space charge buildup in insulators. He is currently Associate Professor at Sorbonne Universités-UPMC since 1991 where he teaches physics, electromagnetism, antennae and high frequency electronics. His research topics include instrumentation for the characterization of insulators, study of sensors and high frequency devices for several domains such as wood industry, automotive... UPMC/ESPCI/CNRS-LPEM/UMR8213, 10 rue Vauquelin, 75005 Paris, France. (email: thierry.ditchi@espci.fr).

Emmanuel Géron was born in Saint-Maurice, France, in 1968. He graduated from the engineering school cole Supérieure de Physique et de Chimie Industrielles of the City of Paris in 1992 and obtained his Ph.D. in Electronics in 1997 from the University of Paris 6, France. He is Associate Professor at ESPCI since 1997. He obtained his Habilitation to lead researches (HDR) in 2013 from the University of Paris 6, Paris. He has made many contributions to Telecommunications and works currently for hyper-frequency circuits and propagation of electromagnetic waves in meta-materials, high resolution space charges systems, and more generally on instrumentation. ESPCI/UPMC/CNRS-LPEM/UMR8213, 10 rue Vauquelin, 75005 Paris, France. (email: emmanuel.geron@espci.fr).

Stéphane Holé born in Pontoise, France, in 1968, studied electronics and instrumentation at Université Pierre et Marie Curie, Paris France. He joined Laboratoire d'Électricité Générale of École Supérieure de Physique et de Chimie Industrielles, Paris France, to study an instrument for measuring fast development of space charges in insulators under rapid voltage variations. It was the topic of his Ph.D. he received in 1996. Currently Professor at Sorbonne Universités-UPMC, he conducted his researches in Laboratoire des Instruments et Systémes d'Ile de France from 1997 to 2007 and leads the Instrumentation Group in Laboratoire de Physique et d'Étude des Materiaux since 2007. His research topics are various such as space charge in insulators and semiconductors (main topic), electrostatic, magnetostatic and ultrasonic sensors. He received the Jack Hollingum Award in 2002 and 2004, and obtained his Habilitation in 2007. He teaches solid state physics, electronics and sensor physics. He is coordinator of the sensors, instrumentation & measurement master at Pierre et Marie Curie University since 2009. His current address is: UPMC/ESPCI/CNRS-LPEM/UMR8213, 10 rue Vauquelin, 75005 Paris, France. (e-mail: stephane.hole@espci.fr).



Yacine Oussar was born in Algiers, Algeria, in 1969. He graduated from the engineering school École Nationale Polytechnique d'Alger in 1993 and obtained his Ph.D. in Machine Learning in 1998 from the University of Paris 6, France. He is Associate Professor at ESPCI since 1998. He obtained his Habilitation to lead researches (HDR) in 2010 from the University of Paris 6, Paris. He has made many contributions to Machine Learning and works currently for electrostatics, electromagnetic waves, partial discharges, and more generally on applications of machine learning techniques for instrumentation. ESPCI/UPMC/CNRS-LPEM/UMR8213, 10 rue Vauquelin, 75005 Paris, France. (email: yacine.oussar@espci.fr).

Maria Merlan was born in Ferrol, Spain, in 1985. She holds a first degree in Telecommunication Engineering and a Master in Radio communication from the Electrical engineering school at the University of Vigo, Galicia 2011. She obtained her Ph.D. in Electronics in 2016 from the University Pierre et Marie Curie in Paris, France. She is specialized in RF antenna and instrumentation and currently working as a research engineer for hyper-frequency technologies in ENS Cachan, 61 Avenue du Président Wilson, (email: maria.merlan@satie.enscachan.fr).

Please cite this article as: H. Daassi-Gnaba, et al., Wood moisture content prediction using feature selection techniques and a kernel method, Neurocomputing (2016), http://dx.doi.org/10.1016/j.neucom.2016.09.005