Sensors and Actuators A 282 (2018) 1–8

Contents lists available at ScienceDirect

Sensors and Actuators A: Physical

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

Application of non-linear statistical tools to a novel microwave dipole antenna moisture soil sensor

A. Berardinelli[a,*], G. Luciani[b,c], M. Crescentini[b,d], A. Romani[b,d], M. Tartagni[b,d], L. Ragni[a,e]

[a] Department of Agricultural and Food Sciences, Alma Mater Studiorum, University of Bologna, Piazza Goidanich 60, 47521 Cesena, FC, Italy
[b] Department of Electrical, Electronic and Information Engineering Ğuglielmo Marconi÷ University of Bologna, Via Venezia 52, Cesena 47521, Italy
[c] Center for Industrial Research on ICT (CIRI ICT), University of Bologna, via Venezia 52, 47521 Cesena, FC, Italy
[d] Advanced Research Centre on Electronic Systems (ARCES), University of Bologna, 40123 Bologna, Italy
[e] Inter-Departmental Centre for Agri-Food Industrial Research, Alma Mater Studiorum, University of Bologna, Piazza Goidanich 60, 47521 Cesena, FC, Italy

ARTICLE INFO

Article history:
Received 20 February 2018
Received in revised form 27 July 2018
Accepted 3 September 2018
Available online 9 September 2018

Keywords:
Dielectric spectroscopy
Soil moisture content
Linear and nonlinear multivariate data analysis
PLS
K-OPLS

ABSTRACT

In this paper we will show the boosting performance of nonlinear machine learning techniques applied to a novel soil moisture sensing approach. A probe consisting in a transmitting and a receiving dipole antenna was set up to indirect assess the moisture content (%) of three different types of soils (silty clay loam, river sand and lightweight expanded clay aggregate, LECA). Gain and phase signals acquired in the 1.0 GHz – 2.7 GHz frequency range were used to built predictive models based on linear PLS regression and on nonlinear Kernel-based orthogonal projections to latent structures (K-OPLS) algorithms. K-OPLS algorithm slightly increased the accuracy of the models built on the gain response on specific kind of soils with respect to classical linear PLS. However, the predictability increases significantly in the case where the models are built starting from a matrix containing all the considered soil samples (silty clay loam + river sand + LECA) achieving R^2 = 0.971 (RMSE = 1.4%) when using K-OPLS non-linear model with respect to R^2 = 0.513 (RMSE = 6.1%) obtained using linear PLS. Therefore, K-OPLS algorithm appears to give a significant improvement to modelling data where nonlinear behaviours occur.

1. Introduction

The spectroscopic and time-domain analyses of the interaction between the electromagnetic wave and the agricultural soil are, at date, widely explored methods for the indirect assessment of its water content [1,2]. The acquired waveforms appeared to contain information related to different soil physico-chemical properties and the quantitative estimation accuracy is affected by two main factors as the used techniques and the statistical tools [3]. Examples of these techniques are Visible, Near and Infrared sensors [4,5], Theta probes, measuring apparent impedance at 100 MHz [6] and the Time Domain Reflectometry (TDR) [7], based on the analysis of the propagation time of the electromagnetic wave through a coaxial cable to a probe immersed in the medium (20 kHz - 1.5 GHz), a function of the soil dielectric permittivity.

Powerful multivariate data analysis tools able to relate two data matrices X (spectra acquired from several samples) and Y (the analytical properties) have played a big role in the development of the techniques [8].

Originated around 1975, the widespread linear multivariate Partial Least Squares (PLS) regression is considered a standard procedure in chemometrics and it has been shown to be potential for extracting useful information starting from highly linearly correlated data coming from bioengineering indirect measurements. The tool uses a two-block quantitative PLS model based on a latent variable decomposition of X and Y variables keeping most of the variance of the explanatory variables. It is well known that PLS regression has proven to be extremely useful in situation where the number of observed variables is much higher than the numbers of acquired samples, typical situation with spectral data [9].

However, non linear behaviours are very frequent in biosystems, such as the light absorbance in milk, dependent on fat content [10], or the dielectric permittivity in microwave region, dependent on the soil moisture [11], just to cite a couple of examples. Samples variability and level of complexity of the matrices together with temperature fluctuations and interactions between sensor and product can negatively affect the robustness of PLS models and cause non linear behaviours as shown in different works conducted on quantitative assessment of fruits chemical properties through

* Corresponding author.
E-mail address: annachi.berardinelli@unibo.it (A. Berardinelli).

https://doi.org/10.1016/j.sna.2018.09.008

Near-infrared spectral measurements [12–14]. Agricultural soil is a complex heterogeneous matrix characterised by organic (humus and different particulate residues) inorganic mineral fractions (proportions of sand, silt and clay particles), moisture and air [15]. Conversely, multivariate regression models based on non linear machine learning tools have shown significant improvements in the accuracy of the prediction of different physical and chemical properties of this complex matrix [5,8,16].

In order to improve robustness of PLS models in presence of non linearity, a considerable number of methods integrating non linear features within the linear PLS algorithm have been proposed. Quadratic PLS [17], smooth bivariate spline function [18], Neural Network PLS [19], Radial Basis Function (RBF) neural networks [20], and Kernel PLS (KPLS) [21] are some examples of the proposed machine learning implementation in PLS modelling. In KPLS the original X variables are transformed into a high-dimensional feature space by a non linear mapping. In this feature space, a linear relationship can be appreciated and the PLS algorithm can then be performed; the feature space is defined after selecting a kernel function providing a similarity measure between pairs of spectra [22]. The accuracy of the KPLS algorithms was tested by analysing images generated by an airborne scanner with nine wavelength bands (from 500 to 10,487 nm) [23], with UV–vis and Fourier Transform Infrared spectra for the prediction of different mixtures contents [24], with NIR spectra for the prediction of apple sugar content [25], and for a rapid screening of water samples containing malathion [26].

Our approach is substantially different with respect to the above-mentioned electro-magnetic techniques. Differently from TDR, it is based on spectra analysis in the frequency domain instead of the time-domain. Then, in contrast with commonly used IR spectra techniques, we perform a spectral analysis of transfer functions involving microwaves. This ensures a better interaction with soil in terms of depth of penetration and also utilizes higher information content given by the phase. Finally, we use non-linear machine learning tools to boost the statistical inference of data.

A new probe in the dielectric sensors panorama characterised by a transmitting and a receiving dipole antenna was set up for the indirect assessment of the moisture content of different types of soils: silty clay loam soil, river sand and lightweight expanded clay aggregate (LECA). This innovative probe requires the previously drilling of the soil and then the insertion of the probe. With respect to traditional TDR probes it could less suffer for incomplete adherence of the soil to the sensor. In fact, the sensing is performed in a large portion of the volume surrounding the probe and any interference, such as air, can be removed by the powerful statistical analysis. Therefore, the information contained in both gain and phase signals acquired in the 1.0 GHz – 2.7 GHz frequency range, will be processed by using the Kernel-based orthogonal projections to latent structures (K-OPLS, an implementation of KPLS with a solution able to separate structured noise). Predictive models of the moisture content will be built starting from data sets characterised by the same soil type or starting from data sets containing all the analysed soil types.

2. Materials and methods

2.1. Probe and acquisition chain

The probe, designed to be inserted in the soil, assembles a transmitting (TX) and a receiving (RX) dipole antenna, spaced 50 mm, located in a 170 mm long PVC sealed pipe, with outer and inner diameter of 16 and 13 mm, respectively. Both TX and RX antennas consists of a ¼ of ring per pole. The dipole was mounted on a nylon ring and placed in the pipe rotated by 90° one with respect

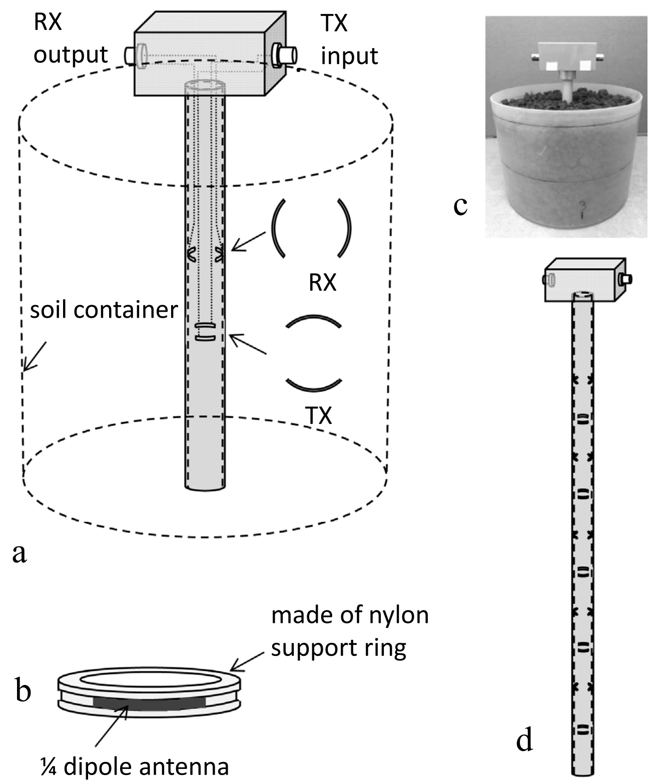

Fig. 1. Layout of the probe containing the ¼ ring dipole antenna (a); a longer probe, containing an array of antennas (b).

to the other, in order to avoid direct coupling of the EM signal from transmitting to receiving antenna. A layout of the probe containing the dipoles is shown in Fig. 1a together with the particulars of the dipole antenna (b) and the probe inserted in the soil (c). The above described prototype was designed for moisture determination in the soil layer pertaining the secondary tillage. A longer probe, containing an array of antennas, suitably spaced, could be constructed for in depth stratified moisture assessment (Fig. 1d). The TX antenna was connected to a sweeper oscillator (HP8350B combined with the HP83592B plug in), by means of a power divider. The signal from the other output of the divider and that coming from the RX antenna were connected to a gain and phase comparator (Analog Devices AD8302) through a 20 dB attenuator. The outputs of the comparator give a measurement of both gain over a ± 30 dB range, scaled to 30 mV/dB, and of phase between signals over a 0°–180° range, scaled to 10 mV/degree. The output of the comparator was connected to a sampling board (National instrument, DAQ USB-4431) with 24 bit of resolution and sampling frequency from 1 kS/s to 102 kS/s. The board was connected to the PC. LabVIEW software was used to display the spectrum and decimate the sampling frequency for reducing the number of data. A layout of the instrumental chain was depicted in Fig. 2. The sinusoidal oscillation (13 dB m) was linearly swept from a frequency of 1.0 GHz to 2.7 GHz in 60 s.

2.2. Soil samples

Waveform acquisition was conducted on three different soil samples: silty clay loam soil (collected from Romagna region agricultural soil, Italy), river sand (Bacchi S.P.A., Italy), and lightweight expanded clay aggregate (LECA) (Laterlite, Italy). According to USDA textural classification [27], the chosen materials are characterised by very distinct physical properties (textural classes). Silty clay loam soil is made of particles with the following size