Soil moisture estimation by using active microwave measurements and support vector regression (SVR)

Thomas Katagis[1], Diego Fernández-Prieto[1], Mattia Marconcini[2], Wouter Dorigo[3]

[1] *European Space Agency, ESA-ESRIN, Frascati, Italy; thomas.katagis@esa.int*
[2] *Deutsches Zentrum für Luft- und Raumfahrt (DLR), Oberpfaffenhofen, Germany*
[3] *Vienna University of Technology, Vienna, Austria*

Abstract. In this study the capabilities of an empirical regression approach for retrieving land surface soil moisture were investigated, by combining time series of ASCAT observations and in-situ data. The Support Vector Regression (SVR) is a novel technique that is utilized for establishing a direct relationship between remote sensing observations and the geophysical parameter of interest. ASCAT series were collocated with the in-situ measurements of soil moisture and the SVR model was trained under different training/testing configuration. The SVR results showed overall satisfactory correlations with the ground-truth data suggesting that the proposed approach is promising. Still, more research should be conducted regarding the model's robustness when applied to different geographical regions, when more features are involved and by focusing on more rigorous data quality control and collocation.

Keywords. Soil moisture, ASCAT, Regression, Support Vector Machines.

1. Introduction

Soil moisture is a key driver that affects the exchange of water, energy and carbon between the land surface and atmosphere through evapotranspiration processes [1], [2]. Accurate spatial-temporal soil moisture measurements are of primary importance for various scientific and operational applications such as numerical weather prediction [3], climate and crop growth modeling [4]-[6], forecast of floods [7]. Traditionally, monitoring of soil moisture dynamics and trends has been conducted with in-situ measurements, while numerous studies have been also based on various hydrological models [8]. Though in-situ datasets offer accurate measurements of soil moisture, their temporal and spatial coverage is limited while long-term field campaigns are quite costly and time-consuming.

Satellite remote sensing offers valuable information about the moisture content of the uppermost soil layer with the advantage of high temporal frequency and extended spatial coverage from local to global scale. In the recent years, data from active and passive microwave sensors have been widely used for soil moisture monitoring at an operational basis. Various global products have become available from sensors such as ERS 1 and 2 [9], [10], AMSR-E [11], [12], WindSat [13], and ASCAT [14]. Additionally, the Soil Moisture and Ocean Salinity (SMOS) sensor of ESA that was launched in 2009, consists a dedicated mission to soil moisture monitoring.

Besides the satellite based retrievals, soil moisture products are also derived from land surface model estimates, as provided by the Global Land Surface Data Assimilation System [15]. For validation purposes, in-situ measurements of soil moisture are considered essential and though the number of ground networks worldwide is limited, the recently introduced International Soil Moisture Network collects and provides harmonized in-situ datasets from many ground stations all over the globe [2].

Various methods and techniques have been proposed for retrieving soil moisture from microwave sensors, including the change detection algorithm for the ERS scatterometer suggested by

Wagner et al. [9], a time series approach proposed by Wen and Su [16], inversion of analytical electromagnetic models [17], artificial neural networks [18]. However, accurate retrievals of soil moisture from satellite sensors is a challenging task, since the electromagnetic signal is strongly affected by soil roughness and heterogeneous vegetated areas [19].

Such limitations can be overcome if long-term time series of data are used, since they account for the soil roughness and vegetation effects and allow detection of the soil moisture content variations [14], [19]. In this study, the capabilities of an empirical regression approach for retrieving soil moisture are investigated, by combining time series of ASCAT observations and in-situ data. The Support Vector Regression (SVR) is a technique that is used for establishing a direct relationship between remote sensing observations and the geophysical parameter of interest [20], [21].

2. Methods

2.1. Datasets

In this study, the training of the SVR algorithm is performed by utilizing suitable in-situ soil moisture data collected in two ground networks in southeast Australia. The in-situ data are derived from the International Soil Moisture Network (ISMN) and refer to top soil layer measurements (0-5cm) for the period 2007-2010. These include measurements from 8 stations of the OzNet Kyeamba and 12 stations of the OzNet Yanco network (Figure 1). The Yanco study area includes mainly agricultural lands of approximately 60x60km and is characterized by gently slopes. The Kyeamba stations are distributed in the Kyeamba Creek catchment, across steep slopes in the upper reaches of the catchment and near the valley mouth and most of the catchment area is subject to grazing and cropping.

The in-situ data have been collocated with the daily ASCAT L1b backscatter measurements, which are resampled data to a fixed Discrete Global Grid (DGG) with 12.5km sampling interval. The active microwave Advanced Scatterometer (ASCAT) is a C-band scatterometer with VV polarization, that uses two sets of three antennas (aft/mid/fore) with an incidence angle range of 250 to 640 and covers a double swath of 550km. ASCAT has been primarily designed to measure the wind speed and its direction over the oceans. However, the measured electromagnetic backscatter is dependent on the dielectric properties of the soil surface layer due to the presence of water, thus providing relatively direct information on soil moisture.

Additionally, soil moisture measurements from the Global Land Data Assimilation System (GLDAS-Noah) as well as from Vienna University of Technology (TU Wien) were used for further evaluation and comparison of the SVR results.

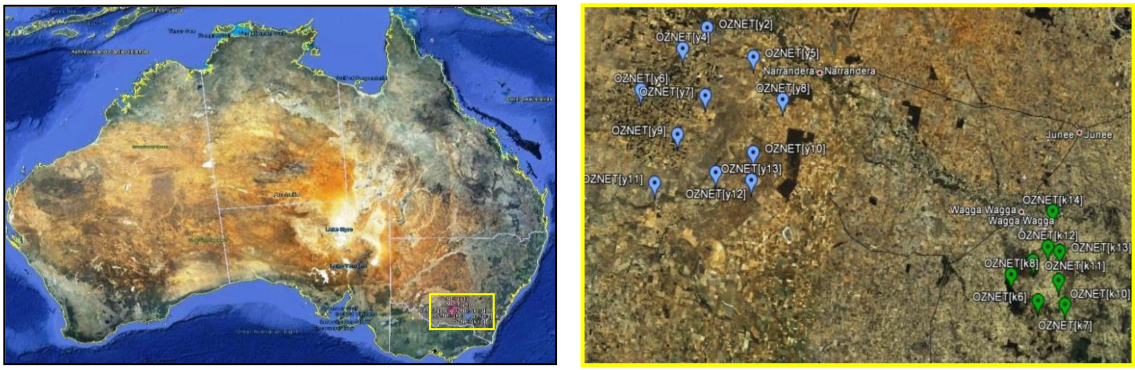

Figure 1. Location of Australian in-situ monitoring stations of soil moisture

2.2. Support Vector Regression background

In order to overcome limitations and drawbacks typical of physical modeling (mainly due to simplification assumptions, the high complexity of the inversion processes and the large number of variables involved), an alternative approach based on empirical modeling is proposed, which relies on regression techniques directly relating remote-sensing observations to the geophysical parameter under analysis by means of suitable interpolation methods. Among the many empirical models proposed so far in the literature, this study focuses on the ε-insensitive support vector regression (ε-SVR) [22], which is one of the support vector machines (SVM) implementation for regression and function approximation. SVR proved capable of high generalization and robustness in a variety of different applications, as well as a limited complexity in handling the learning phase. In particular, instead of minimizing the observed training error, SVR attempts to minimize the generalization error bound so as to achieve generalized performance.

In order to retrieve the geophysical parameter under analysis y (i.e., the in-situ soil moisture) given a set $\mathbf{x}=\{x^1,\dots,x^M\}$ of M different variables of interest (e.g. the ASCAT observations) the goal of SVR is to find a smooth approximation $\tilde{f}$ of the true function f describing the desired input-output mapping, i.e., $y=f(\mathbf{x})+e$ (where e is a Gaussian random variable with zero mean and unit variance accounting for all the noise contributions affecting the estimation problem). In particular, given N training samples $\mathrm{T}=\{(\mathbf{x}_i,y_i)\}_{i=1}^N$ composed by pairs of *in situ* measurements y_i for the geophysical parameter y and the corresponding set of remote-sensing observations $\mathbf{x}_i=\{x_i^1,\dots,x_i^M\}$ for the M variables of interest (acquired at the same location of y_i), the ε-SVR aims at determining $\tilde{f}$ such that the estimated targets $\hat{y}_i$ has at most ε deviation from the training data (this allows to define an ε-insensitive tube surrounding $\tilde{f}$, hence increasing the robustness to noise and small errors in T). ε-SVR first maps the input data to a higher dimensional (possibly infinite) kernel feature space by means of a non-linear mapping $\Phi(\cdot)$ and then solves the linear mapping there, i.e.

$$\hat{y}_i=\tilde{f}(\mathbf{x}_i)=\Phi(\mathbf{x}_i)\cdot\mathbf{w}+b \tag{1}$$

where $\mathbf{w}$ is a weight vector in the feature space and b is the bias term in the regression. In the SVR formulation the optimal $\mathbf{w}$ and b are determined solving the following constrained minimization problem [22]:

$$\begin{cases}\min\limits_{\mathbf{w},b}\left\{\frac{1}{2}\|\mathbf{w}\|^2+C\sum_{i=1}^{N}(\xi_i+\xi_i^*)\right\}\\ y_i-\mathbf{w}\cdot\Phi(\mathbf{x}_i)-b \quad\le\quad \varepsilon+\xi_i\\ \mathbf{w}\cdot\Phi(\mathbf{x}_i)+b-y_i \quad\le\quad \varepsilon+\xi_i\\ \xi_i,\xi_i^*\ge 0 \quad \forall i=1,\dots,N\end{cases} \tag{2}$$

where ξ_i and ξ_i^* are slack variables measuring the distance of the training samples lying outside the ε-insensitive tube from the tube itself, whereas the parameter $C\in\mathbb{R}_0^+$ determines the trade-off between the flatness of $\tilde{f}$ and the amount up to which deviations larger than ε are tolerated. It is possible to demonstrate that the problem in (2) can be written in its dual formulation by means of the Lagrange optimization theory, hence leading to the following convex problem (which has a unique solution corresponding to the global maximum of the cost function):

$$\begin{cases} \max\limits_{\boldsymbol{\alpha},\boldsymbol{\alpha}^*} \begin{Bmatrix} \sum_{i=1}^{N} y_i(\alpha_i - \alpha_i^*) - \varepsilon \sum_{i=1}^{N} (\alpha_i + \alpha_i^*) + \\ -\frac{1}{2} \sum_{i,j=1}^{N,N} (\alpha_i - \alpha_i^*)(\alpha_j - \alpha_j^*) k(\mathbf{x}_i, \mathbf{x}_j) \end{Bmatrix} \\ \sum_{i=1}^{N} (\alpha_i - \alpha_i^*) = 0 \\ 0 \le \alpha_i \le C \quad \forall i = 1, ..., N \\ 0 \le \alpha_i^* \le C \quad \forall i = 1, ..., N \end{cases} \tag{3}$$

where α_i and α_i^* are the Lagrange multipliers associated with training samples, and $k(\cdot,\cdot)$ is a kernel function satisfying the Mercer's condition [23] such that $k(\cdot,\cdot) = \Phi(\cdot) \cdot \Phi(\cdot)$. The employment of a kernel function is of great importance as it allows to evaluate the similarity between pairs of samples in the transformed space as a function of the samples in the input space, thus avoiding to explicitly define $\Phi(\cdot)$ while reducing the complexity of the mapping issue. The optimization of (3) constitutes a quadratic programming (QP) problem and it is possible to prove that the final solution is given by:

$$\hat{y} = \tilde{f}(\mathbf{x}) = \sum_{i=1}^{N} (\alpha_i - \alpha_i^*) k(\mathbf{x}_i, \mathbf{x}_j) + b \tag{4}$$

The experiments of this study have been performed using a Gaussian kernel and the SVR model selection was conducted by using a 5-fold cross-validation strategy.

3. Results

The variables used in the SVR experiments included both the original L1b ASCAT backscatter measurements (sig a/m/f), and the normalized ASCAT measurements (sig40). The latter are generated by extrapolating the original measurements to a reference incidence angle of 400 and are corrected for the seasonal influence of vegetation. The daily sensor measurements were then collocated with the in-situ soil moisture data in the learning phase of the SVR under various training/testing configurations. The total number of collocated samples was 8078 for the OzNet Kyeamba and 12530 for the OzNet Yanco network.

Table 1. SVR regression/correlation performances for the various configurations for the two networks (*doy* refers to day of the year)

	Training	Testing	Features	vs in-situ					
				SVR r^2	**SVR r**	**GLDAS r^2**	**GLDAS r**	**TUW r^2**	**TUW r**
OZNET K	2007-08	2009	sig40,doy	0.585	0.765	0.495	0.704	0.536	0.732
	2007-08	2009-10	sig40,doy	0.500	0.707	0.518	0.720	0.535	0.731
	2007-08	2009	siga/m/f,doy	0.542	0.736	0.495	0.704	0.536	0.732
	2007-08	2009-10	siga/m/f doy	0.440	0.664	0.518	0.720	0.535	0.731
OZNETY	2007-08	2009	sig40,doy	0.481	0.694	0.438	0.662	0.466	0.683
	2007-08	2009-10	sig40,doy	0.473	0.688	0.628	0.792	0.616	0.785
	2007-09	2010	sig40,doy	0.348	0.590	0.622	0.788	0.587	0.766
	2007-08	2009	siga/m/f,doy	0.483	0.695	0.438	0.662	0.466	0.683
	2007-08	2009-10	siga/m/f,doy	0.467	0.683	0.628	0.792	0.616	0.785
	2007-09	2010	siga/m/f,doy	0.453	0.673	0.622	0.788	0.587	0.766

Table 1 shows the correlation coefficient r and the square correlation coefficient r2 obtained for the experiments carried out over the selected period in various training and testing configurations for the two networks. Besides the SVR results, correlations of the GLDAS and the TU Wien soil moisture estimations versus the in-situ data are also presented. The SVR outcome for the OzNet

Kyeamba network reveals overall good correlations, higher than the OzNet Yanco network. When compared with the GLDAS and TU Wien performance, SVR correlations are in good agreement or higher for certain configurations. Furthermore, when the normalized ASCAT (sig40) are employed instead of the original measurements a minor improvement of the correlations is achieved.

Regarding the OzNet Yanco network, SVR performance is lower overall compared to the to the other soil moisture products except for the 2007-08/2009 training/testing configuration. Unlike the previous results, employment of the sig40 variable did not improve the correlations, indicating that in-situ data quality or proper data collocation could be affecting the SVR performance more than the incidence angle dependency of the backscatter measurements. Consequently, a more thorough investigation should be further performed focusing on the usage of additional features and addressing the ground-truth spatial sampling errors. Still, the regression model achieves satisfactory results by using only the ASCAT time series as a feature and given the fact that a direct comparison of the coarse resolution satellite data with the sparse in-situ soil moisture measurements remains a challenging task.

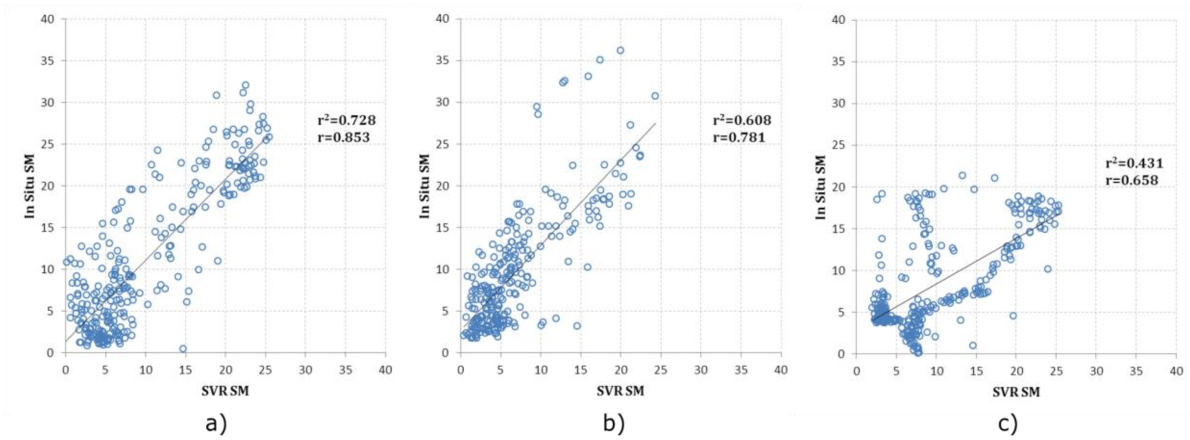

Figure 2. SVR results obtained for individual network stations. a) and b) plots exhibit results for OzNet Kyeamba stations and c) exhibits results for an OzNet Yanco station. The training period is 2007-2008 and testing period 2009, with sig40 variable as input.

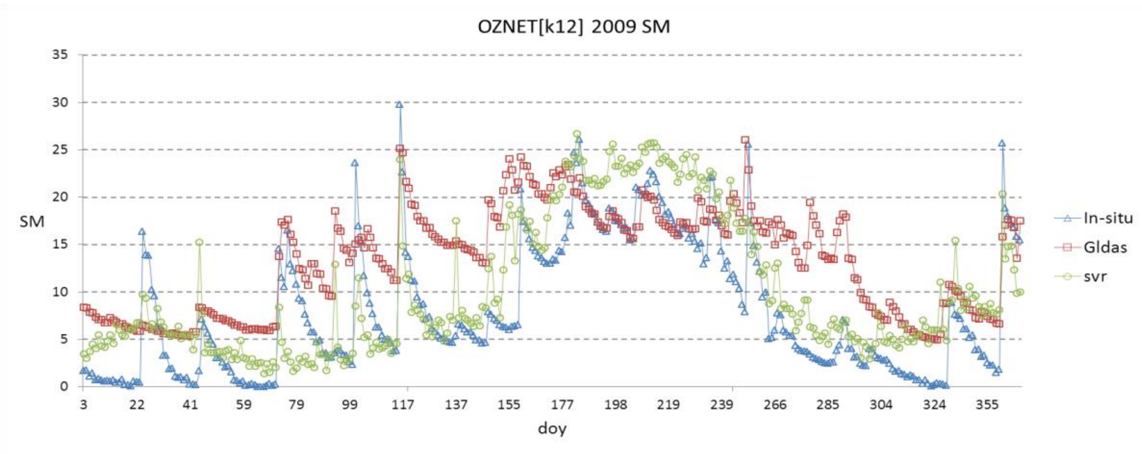

Figure 3. In-situ, GLDAS and SVR SM values for an individual network station. The testing period is 2009, with sig40 feature as input.

In Figure 2 scatterplots of SVR results versus the in-situ data are presented for individual network stations, showing how performances vary for each site within the relatively homogeneous study areas. In addition, soil moisture values are plotted in time for an individual station exhibiting the agreement among the SVR, in-situ and GLDAS estimations (Figure 3). Finally, a cross-network validation was also performed in order to investigate the SVR performance when training and testing the model in different geographical areas. Therefore, all OzNet Kyeamba samples were used to train the SVR model and then test it on OzNet Yanco dataset, resulting in a satisfactory correlation of 0.653.

4. Conclusions

In this study, the potential of Support Vector Regression (SVR) for retrieving accurate measurements of land surface soil moisture from satellite observations was investigated. The SVR technique was implemented after collocation of ASCAT observations with in-situ soil moisture measurements, exhibiting low computation time and reduced complexity during the training phase. Estimated SVR soil moisture values are in general well correlated with in-situ data and results exhibit minor variations depending on the different training/testing configurations. Despite the comparison of in-situ points with the coarse resolution satellite series, the SVR succeeded in identifying the temporal correlation between the measurements. Employment of the ASCAT normalized (sig40) instead of the original backscatter measurements revealed only minor improvements of correlations, posing the need for a more detailed investigation. Therefore, the scope of the study will be extended by implementing the SVR method in different geographical areas, by employing additional features and by applying an even more rigorous data quality control and collocation.

References

[1] Dorigo, W., De Jeu, R., Chung, D., Parinussa, R, Liu, Y., Wagner, W., Fernández-Prieto, D., 2012. *Evaluating global trends (1988–2010) in harmonized multi-satellite surface soil moisture*. Geophysical Research Letters, 39, L18405

[2] Jung, M. *et al*., 2010. *Recent decline in the global land evapotranspiration trend due to limited moisture supply*. Nature, 467(7318), 951–954.

[3] Albergel, C., Calvet. J. C., De Rosnay, P., Balsamo, G., Wagner, W., Hasenauer. S., Naeimi, V., Martin, E., Bazile, E., Bouyssel, F., & Mahfouf, J. F., 2010. *Cross-evaluation of modelled and remotely sensed surface soilmoisture with in situ data in Southwestern France*. Hydrology and Earth System Sciences, 14, 2177–2191.

[4] Diermeyer, P. A., 2012. *Using a global soil wetness dataset to improve seasonal climate simulation*. Journal of Climate, 13, 2900–2922.

[5] Georgakakos, K. P. & Carpenter, M., 2006. *Potential value of operationally available and spatially distributed ensemble soil water estimates for agriculture*, Journal of Hydrology, 328, 177–191, 2006.

[6] De Wit, A. J. & van Diepen, A. J., 2007. *Crop model data assimilation with the ensemble Kalman filter for improving regional crop yield forecasts*. Agricultural and Forest Meteorology, 146, 38–56.

[7] Brocca, L., Melone, F., Moramarco, T., Wagner, W., Naeimi, V., Bartalis, Z. & Hasenauer S, 2010. *Improving runoff prediction through the assimilation of the ASCAT soil moisture product*. Hydrology and Earth System Sciences, 14, 1881–1893.

[8] Sheffield, J. & E. F. Wood, 2008. *Global trends and variability in soil moisture and drought characteristics, 1950–2000, from observation-driven simulations of the terrestrial hydrologic cycle*. Journal of Climate, 21(3), 432–458.

[9] Wagner, W., Lemoine, G. & Rott, H., 1999. *A method for estimating soil moisture from ERS scatterometer and soil data*. Remote Sensing of Environment, 70, 191–207.

[10] Scipal, K., Wagner, W., Trommler, M. & Naumann, K., 2002. *The global soil moisture archive 1992–2000 from ERS scatterometer data: First results*. IGARSS 2002: IEEE International Geoscience and Remote Sensing Symposium and 24th Canadian Symposium on Remote Sensing, Toronto, Canada,

[11] Njoku, E. G., Jackson, T. J., Lakshmi, V., Chan, T. K. & Nghiem, S. V., 2003. *Soil moisture retrieval from AMSR-E*. IEEE Transactions on Geoscience and Remote Sensing, 41, 215–229.

[12] Owe, M., de Jeu, R. & Holmes, T., 2008. *Multisensor historical climatology of satellite-derived global land surface moisture*. Journal of Geophysical Research, 113, F01002, doi:10.1029/2007jf000769.

[13] Li, L., Gaiser, P. W., Gao, B. C., Bevilacqua, R. M., Jackson, T. J., Njoku, E. G., Rudiger, C., Calvet, J. C. & Bindlish, R., 2010. *WindSat global soil moisture retrieval and validation*. IEEE Transactions on Geoscience and Remote Sensing, 48, 2224–2241.
[14] Naeimi, V., Scipal, K., Bartalis, Z., Hasenauer, S. & Wagner, W., 2009. *An Improved Soil Moisture Retrieval Algorithm for ERS and METOP Scatterometer Observations*. IEEE Transactions on Geoscience and Remote Sensing, 47, 1999-2013.
[15] Rodell, M., Houser, P. R., Jambor, U., Gottschalck, J., Mitchell, K., Meng, C. J., Arsenault, K., Cosgrove, B., Radakovich, J., Bosilovich, M., Entin, J. K., Walker, J. P., Lohmann, D. & Toll, D., 2004. *The Global Land Data Assimilation System*. Bulletin of the American Meteorological Society, 85, 381–394.
[16] Wen, J. & Z., Su, 2003. *A time series based method for estimating relative soil moisture with ERS wind scatterometer data*. Geophysical Research Letters, 30(7), 1397.
[17] Y., Oh, Sarabandi, K., & F. T., Ulaby, 2002. *Semi-empirical model of the ensemble-averaged differential Mueller matrix for microwave backscattering from bare soil surfaces*. IEEE Transactions on Geoscience and Remote Sensing, 40(6), 1348–1355.
[18] Weimann, A., 1998. *Inverting a microwave backscattering model by the use of a neural network for the estimation of soil moisture*. Geoscience and Remote Sensing Symposium Proceedings, IGARSS '98, IEEE International , vol.4, 1837-1839.
[19] Paloscia, S., Pampaloni, P., Pettinato, S. & Santi, E., 2008. *A comparison of algorithms for retrieving soil moisture from ENVISAT/ASAR Images*. IEEE Transactions on Geoscience and Remote Sensing, 46(10), 3274–3284.
[20] Pasolli, L., Notarnicola, C. & Bruzzone, L., 2011. *Estimating soil moisture with the support vector regression technique*. IEEE Geoscience and Remote Sensing Letters, 8(6), 1080–1084.
[21] Sabia, R., Marconcini, M., Katagis, T., Fernandez-Prieto, D., Martinez, J. & Portabella, M., 2012. *Preliminary results of SMOS salinity retrieval by using Support Vector Regression (SVR).* Geoscience and Remote Sensing Symposium (IGARSS), 2012 IEEE International , 2629-2632,.
[22] Vapnik, V., 1995.*The Nature of Statistical Learning Theory*. New York: Springer-Verlag,
[23] Courant, R. & D. Hilbert, 1953. *Methods of Mathematical Physics*. New York: Wiley.