Krzysztof Lamorski[1], Cezary Sławiński[1], Felix Moreno[2], Wojciech Skierucha[1], José L. Arrue[3]
[1]Institute of Agrophysics, Polish Academy of Sciences, Lublin, Poland
[2]Instituto de Recursos Naturales y Agrobiología de Sevilla, IRNAS-CSIC, Seville, Spain
[3]Estación Experimental Aula Dei, EEAD-CSIC, Zaragoza, Spain

Introduction

Soil hydrologic parameters have a great importance for soil water transport processes occurring in the soil medium. Retention curve, which describes relation between soil water content and soil water potential is especially important, as it determines soil water availability for plants.

Measurements give strict evaluation of hydraulic properties of soils. Unfortunately measurement of soil water retention curve is highly time consuming and requires specialized equipment.

This work presents point pedotransfer function models of the soil water retention curve. Developed models allow for estimation of the soil water content based on following soil characteristics: soil granulometric composition, total porosity and bulk density. Soil water content is evaluated for the specified soil water potentials: -0.98 kPa, -3.10 kPa, -9.81 kPa, - 31.02 kPa, -491.66 kPa and -1554.78 kPa.

Support Vector Machines (SVM) methodology was used for model development. Alternative to previous attempts ν-SVM method was used for models development and results compared with C-SVM based models. In the work two different types of the SVM kernel function was used and results compared: radial basis kernel function with the linear one.

Soil properties used as input variables for the PTF models are: sand fraction, clay fraction, total porosity and bulk density. Models approximate value of the soil water content for seven fixed values of the soil water potential.

Soil dataset

Two soil datasets were used in this study. The main dataset: extract from the Soil Profiles Bank of Polish Mineral Soils database, 810 soil samples (290 different soil profiles). The second dataset: Spanish soils, 134 samples (taken from eight soil profiles).

The following soil parameters were collected: soil water content for various seven soil water potential values: -0.10 kPa, -0.98 kPa, -3.10 kPa, -9.81 kPa, - 31.02 kPa, -491.66 kPa and – 1554.78 kPa, particle size distribution, total porosity, bulk density.

The main soil database of Polish soils was randomly split into two subsets: training dataset (565 samples) and testing dataset (primary testing dataset, 245 samples).

Spanish soils dataset was used as an additional testing dataset (secondary testing dataset) for model validation purposes.

SVM models were build using training dataset, and tested against test datasets. The K-fold cross-validation technique was used for model elaboration.

Support Vector Machines and model formulation

Support Vector Machines is one of the class of soft-computing techniques. Originally SVM was developed for solving classification problems, then its usage was extended to regression type applications for function estimation. SVM methodology uses training datasets for model building.

Linear kernel function:

$$K\left(\bar{x}^i, \bar{x}^j\right) = \bar{x}^i \cdot \bar{x}^j$$

Radial basis kernel function:

$$K\left(\bar{x}^i, \bar{x}^j\right) = e^{-\gamma\left|\bar{x}^i - \bar{x}^j\right|^2}$$

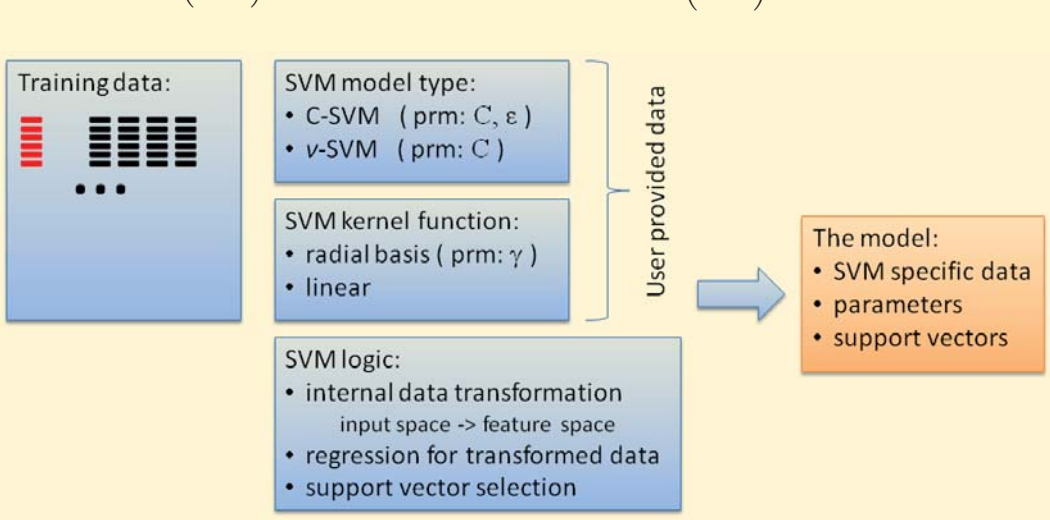

Figure 1: The idea of the SVM computing

SVM models depend on parameters which have to be adjusted, see Table 1. There are following parameters: C – SVM cost parameter, ϵ – "tube" width parameter and γ parameter of radial basis kernel function. The number of SVM methodology connected model parameters was dependent on the model type:

Model name	SVM method	kernel function	model parameters	nr. of prm.
C-radial	C-SVM	radial	C, ϵ, γ	3
C-linear	C-SVM	linear	C, ϵ	2
nu-radial	ν-SVM	radial	C, γ	2
nu-linear	ν-SVM	linear	C	1

Table 1: The specification of tested models and it's parameters

GA model parameters optimisation

Genetic algorithms were used for model parameters selection. The aim function used for model parameters search was simply the RMSE for linear kernel function based models. When RMSE was used as the aim function for radial basis kernel based SVM models then GA optimisation process lead to strong overfitting of the model.

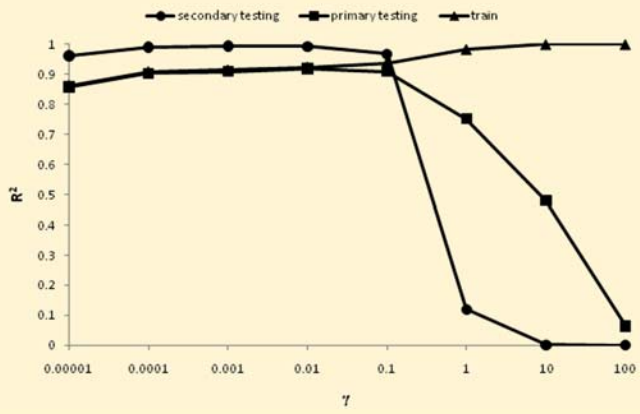

Figure 2: Example exposition of the overfitting phenomena

Due to the overfittng phenomenon, for the radial basis kernel function based SVM model's parameter search procedure, alternative form of the aim function was proposed:

$$F_{aim}(RMSE, nSV) = 1 - e^{-\frac{RMSE^2}{2\sigma^2_{train}}} e^{-\frac{(nSV - nSV_{exp})^2}{a}}$$

where:

- RMSE - root mean square error,
- nSV - number of support vectors in the elaborated model,
- nSV_{exp} is a number of support vectors expected in the developed model,
- a and σ^2_{train} are constants connected with the slopes and the widths of the function.

Results

Below are summarized performance indicators of elaborated models: root mean square error (RMSE) and coefficient of determination (R^2) for primary testing dataset, secondary testing dataset and the training dataset. In case of the number of support vectors, RMSE and R2 for the training dataset, presented values are averages other ten submodels, values in brackets are standard deviations.

pot. [kPa]	model	nr. of SV	training dataset		prim. testing		sec. testing	
			RMSE	R^2	RMSE	R^2	RMSE	R^2
-0.98	C-linear	152.6 (28.94)	0.0185 (0.0004)	0.91 (0.0053)	0.0190	0.91	0.0090	0.99
	C-radial	243.0 (0.00)	0.0124 (0.0008)	0.96 (0.0049)	0.0247	0.84	0.0715	0.28
	nu-linear	248.2 (0.42)	0.0186 (0.0004)	0.91 (0.0053)	0.0190	0.91	0.0068	0.99
	nu-radial	253.2 (4.59)	0.0172 (0.0005)	0.92 (0.0052)	0.0174	0.92	0.0077	0.99
-3.1	C-linear	110.7 (13.57)	0.0432 (0.0007)	0.68 (0.0068)	0.0384	0.69	0.0300	0.93
	C-radial	242.9 (0.57)	0.0260 (0.0029)	0.88 (0.0298)	0.0499	0.52	0.0841	0.18
	nu-linear	248.3 (0.82)	0.0440 (0.0007)	0.67 (0.0067)	0.0375	0.69	0.0245	0.95
	nu-radial	255.6 (9.05)	0.0397 (0.0027)	0.73 (0.035)	0.0354	0.72	0.0191	0.97
-9.81	C-linear	136.0 (12.38)	0.0442 (0.0005)	0.82 (0.0029)	0.0446	0.78	0.0391	0.89
	C-radial	242.9 (0.57)	0.0260 (0.0007)	0.94 (0.004)	0.0487	0.74	0.1102	0.02
	nu-linear	248.9 (0.74)	0.0443 (0.0005)	0.82 (0.003)	0.0451	0.78	0.0387	0.90
	nu-radial	254.3 (6.72)	0.0378 (0.0024)	0.87 (0.017)	0.0369	0.85	0.0368	0.88
-31.02	C-linear	144.4 (33.68)	0.0455 (0.0005)	0.82 (0.0031)	0.0492	0.75	0.0495	0.89
	C-radial	243.1 (0.57)	0.0273 (0.0007)	0.94 (0.0036)	0.0532	0.72	0.1228	0.06
	nu-linear	248.1 (0.74)	0.0459 (0.0006)	0.82 (0.0034)	0.0496	0.75	0.0487	0.89
	nu-radial	252.3 (2.83)	0.0407 (0.0014)	0.86 (0.0096)	0.0440	0.80	0.0469	0.86
-491.66	C-linear	185.8 (44.52)	0.0466 (0.0009)	0.75 (0.0058)	0.0478	0.69	0.0849	0.71
	C-radial	243.2 (0.63)	0.0268 (0.0008)	0.92 (0.0058)	0.0589	0.59	0.1021	0.33
	nu-linear	249.4 (1.43)	0.0467 (0.0009)	0.75 (0.0061)	0.0479	0.69	0.0874	0.71
	nu-radial	252.1 (4.28)	0.0435 (0.0019)	0.78 (0.0181)	0.0457	0.72	0.0802	0.71
-1554.78	C-linear	172.0 (18.22)	0.0450 (0.0011)	0.69 (0.0104)	0.0459	0.63	0.0601	0.73
	C-radial	243.3 (0.48)	0.0266 (0.001)	0.89 (0.0077)	0.0590	0.49	0.1077	0.27
	nu-linear	248.4 (0.52)	0.0452 (0.0012)	0.69 (0.0108)	0.0458	0.63	0.0620	0.73
	nu-radial	254.2 (5.43)	0.0423 (0.0019)	0.73 (0.0208)	0.0439	0.66	0.0605	0.71

Table 2: Performance comparison of the elaborated PTF models

Conclusions

- Elaborated models showed good agreement with measured soil water retention data. Achieved coefficient of determination values were from 0.92 to 0.66. Studies confirmed usability of the ν-SVM methodology and the linear kernel function for pedotransfer function developments.
- The SVM methodology was successfully applied for the water retention modelling. Developed models showed good agreement with measured data.
- The ν-SVM method is suitable for the developments of PTF models for retention curve approximation. The advantage of using this method is a limited number of model parameters.
- The ν-SVM based models with fixed value of nu=0.5 parameter, showed performance better or the same as C-SVM based models with adjustable epsilon parameter.
- Investigated linear kernel function may be used successfully instead of the radial basis one, for point PTF developments.
- Proposed form of the aim function is useful for optimal parameters determination for the radial basis kernel function based ν-SVM models, and allows for selection of appropriate number of support vectors.

http://www.ipan.lublin.pl k.lamorski@ipan.lublin.pl