Soft-computing techniques in soil hydrological parameters modelling

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

Agrophysics for Quality of Life
31st May - 3rd June 2011
Lublin, Poland

The idea of the soft-computing

Soft-computing models:

- based on the data learning,
- does not provide analytical solution of the problem,solutions are by by definition inexact and approximate,
- allow for modelling of the properties or behaviour of the complex systems without deep insight,
- gives specific solution for currently modelled phenomenon.

Classical models:

- based on full physical-mathematical modelling,
- described by some class of exact mathematical equations,
- often expensive in the sense resources utilised for solving the problem (computing time),
- universal for the given phenomenon described.

Typical soft-computing methods

Mathematical methods considered as the soft-computing methods include:

- ANN - Artificial Neural Networks,
- kNN - k-Nearest Neighbour,
- Support Vector Machines,
- Fuzzy Logic,
- Genetic Algorithms.

Hydrological pedotransfer functions

Hydrological pedotransfer functions - allow for approximation of the soil hydrological parameters based on some basic soil characteristics.

Soil hydrological parameters:

- soil water retention curve,
  - soil water field capacity,
  - wilting point,
- soil water hydraulic conductivity,
  - saturated hydraulic conductivity,
  - unsaturated hydraulic conductivity.

The aim

To develop SVM based pedotransfer function which will allow for evaluation of soil water - soil water potential relationship.

To limit number of the SVM model parameters, needed for model formulation.

Soil material

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 – pF 0, -0.98 kPa – pF 1, -3.10 kPa – pF 1.5, -9.81 kPa – pF 2, - 31.02 kPa – pF 2.5, -491.66 kPa – pF 3.7 and – 1554.78 kPa – pF 4.2,
- particle size distribution,
- total porosity,
- bulk density.

Working datasets

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

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

SVM models were build using training dataset, and tested against test datasets.

Secondary testing dataset was used for additional model performance verification, for soils with completely different origin.

The K-fold cross-validation technique was used for model elaboration.

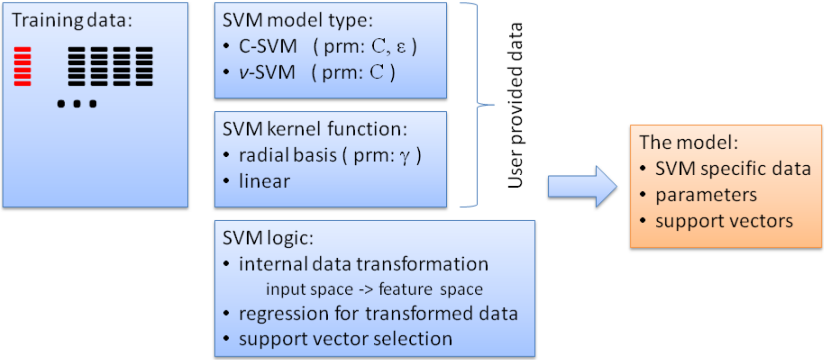
SVM - methodology

Linear kernel function:

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

Radial basis kernel function:

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



Models type selection

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

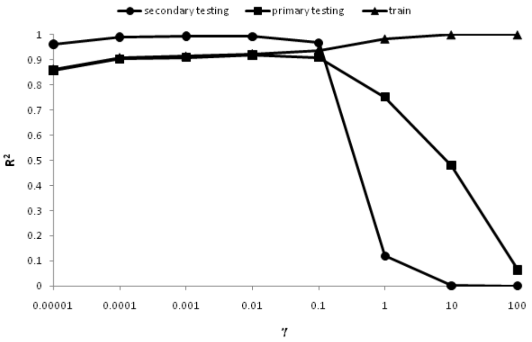
Parameters of the PTF models are:

- sand fraction – which has high correlation for middle values of soil water potentials,
- clay fraction – which has high correlation with water content for high values of soil water potential,
- total porosity - very high correlation for low values of soil water potentials,
- bulk density – high correlation for low values of the soil water potentials.

Models parameters selection

Genetic algorithms were used for model parameters selection.

Initially the aim function was simply RMSE, but GA optimisation process lead to strong overfitting for radial basis kernel based SVM models.



Alternative form of the aim function

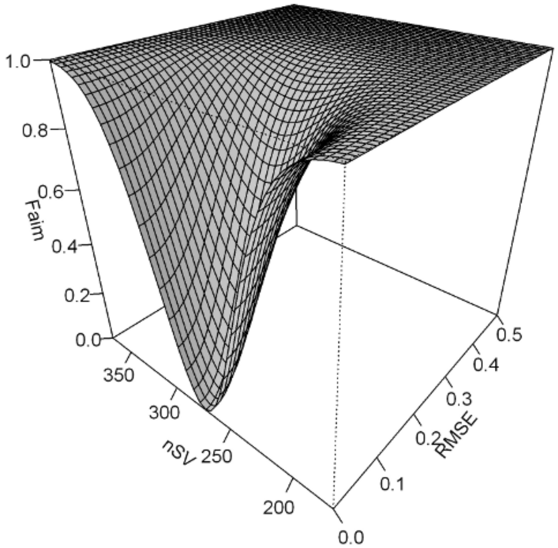
Due to overfittng phenomenon, alternative form of the aim function was proposed instead of RMSE:

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

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.

Alternative form of the aim function ...



Results

pot. [kPa]	model	nr. of SV	training dataset RMSE	training dataset R^2	prim. testing RMSE	prim. testing R^2	sec. testing RMSE	sec. testing 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

Conclusions

- The SVM methodology was successfully applied for the water retention modelling.
- Developed models showed good agreement with measured data for the soils with the same origin (Polish soils, temperate climate) and with the soils with the different origin (Spanish soils, mediterranean climate).
- 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, showed performance better or the same as C-SVM based models.

The end

Thank you for your attention