Available online at www.sciencedirect.com

SCIENCE DIRECT®

ELSEVIER

Environmental Modelling & Software ■■ (2005) ■■■–■■■

Environmental Modelling & Software

www.elsevier.com/locate/envsoft

PTFIndicator: An IRENE_DLL-based application to evaluate estimates from pedotransfer functions by integrated indices

G. Fila*, M. Donatelli, G. Bellocchi

Research Institute for Industrial Crops, via di Corticella 133, 40128 Bologna, Italy

Received 7 April 2004; received in revised form 24 November 2004; accepted 28 November 2004

Abstract

Pedotransfer functions (PTFs) are physical–mathematical models which allow estimating soil hydraulic properties from soil data. In the absence of direct measurements, they represent a valuable tool for providing agro-ecological models with soil input estimations, whose accuracy is currently a key issue, especially in field- and regional-scale modelling which rely on soil information databases. The class library IRENE_DLL (Integrated Resources for Evaluating Numerical Estimates – Dynamic Link Library) is a repository of statistical methods and procedures designed to properly handle comparisons between model estimates and actual measurements. A distinct feature is the support for developing integrated evaluation indices based on a two-stage, fuzzy-based approach. We implemented this approach into PTFIndicator, a Microsoft Excel application based on IRENE_DLL, to evaluate the performance of PTFs. This paper introduces the overall functioning of the application, and briefly discusses one relevant worked example.

Keywords: Component object model; Dynamic link library; Fuzzy logic; IRENE_DLL; Pedotransfer function; PTFIndicator; MS Excel; Visual basic for applications

Software availability

Name of software: PTFIndicator
Developer: Gianni Fila
Contact address: ISCI, Bologna, Italy
Telephone: +39-051-6316811
Fax: +39-051-374857
Email: g.fila@isci.it
Availability: http://www.sipeaa.it/tools

1. Introduction

Pedotransfer functions (PTFs) are relationships between soil variables to estimate soil properties from existing soil data sets (Bouma, 1989; Hamblin, 1991). Given the cost and difficulty of measuring soil hydraulic properties in the field, PTFs represent a valuable alternative for providing agro-ecological models with soil input estimations from readily-available soil data (Wösten, 2002). Hence, evaluation of PTFs accuracy is a key issue in field- and regional-scale modelling which rely on soil information databases (Tietje and Tapkenhinrichs, 1993; Tietje and Hennings, 1996; Hennings et al., 1997; Pachepsky et al., 1999). Although not exhaustive in every aspect of model capability, the square error statistics are widely used to evaluate PTFs accuracy (e.g., Wösten et al., 2001). A review of performance measures for possible use in PTFs evaluation has been given by Imam et al. (1999), who emphasized that no single statistic can adequately describe PTFs performance.

IRENE_DLL (Integrated Resources for Evaluating Numerical Estimates – Dynamic Link Library, Fila et al.,

* Corresponding author. Tel.: +39 051 6316 845; fax: +39 051 3748 857.
E-mail address: g.fila@isci.it (G. Fila).

1364-8152/$ - see front matter
doi:10.1016/j.envsoft.2005.01.001

2003) is a dynamic link library implemented using COM technology (Microsoft Component Object Model, http://www.microsoft.com/com), that incorporates the most relevant statistical procedures for use in model evaluation. This library was developed as part of a larger action towards the establishment of a commonly agreed architecture for agronomic and agro-meteorological modelling (Donatelli et al., 2002; He et al., 2002). The methods provided by IRENE_DLL allow developers to extend existing skills and code sets for the rapid development of statistical evaluation routines within modelling projects, and to increase the effectiveness of analysis by creating customized Windows-based applications.

In the context of PTFs evaluation we have implemented an approach which exploits the features of IRENE_DLL within a Microsoft Excel-based client application. The aim of this paper is to illustrate this application and its applicability to real cases.

2. Evaluation of PTFs by integrated indices

Donatelli et al. (in press-b) developed two fuzzy-based integrated indices (indicators), namely I_{PTFSW} and I_{PTFRC}. The first is used to evaluate PTFs which estimate soil water content at critical points (e.g., at field capacity or at wilting point), while the second is used for the evaluation of PTFs which provide estimates of soil water content over a range of soil water potentials (values from soil water retention curves). The rationale of both indicators is based on the methodology proposed by Bellocchi et al. (2002), where three issues were pointed out as relevant in a model quality judgment: (i) ability of the model to produce small residuals; (ii) high correlation between estimates and measurements; (iii) absence of systematic patterns in model residuals. Based on such considerations, we defined three modules to match the aforementioned criteria, namely Accuracy, Correlation and Pattern. A *module*, in the terminology adopted by IRENE_DLL, is an evaluation index calculated via a fuzzy-based procedure from one or more basic statistics (e.g., RRMSE, relative mean square error). For each module, a dimensionless value between 0 (best model response) and 1 (worst model response) is calculated. The Sugeno method of fuzzy inference was adopted (Sugeno, 1985). Three membership classes (or subsets) are defined for all indices, according to an expert judgment, namely favourable (F), unfavourable (U) and partial (or fuzzy) membership, using S-shaped curves as transition probabilities in the range F to U. A two-stage design of fuzzy-based rules inferring system is applied (Table 1): first, several indices are aggregated into modules and then, using the same procedure, the modules are aggregated in a second level integrated index (again, ranging from 0 to 1), which we called *indicator*.

Table 1
The modules Accuracy, Correlation, and Pattern and their inputs for two indicators

Input[a]	Accuracy	Correlation	Pattern
RRMSE ($F \le 30$; $U \ge 60$)	0.250		
EF ($F \ge 0.5$; $U \le 0.0$)	0.250		
r ($F \ge 0.9$; $U \le 0.7$)		0.150	
$PI_{D_{50}}$ ($F \le 0.03$; $U \ge 0.06$) (b: I_{PTFSW}; c: I_{PTFRC})			0.175[b] 0.105[c]
PI_{OC} ($F \le 0.05$; $U \ge 0.08$) (b: I_{PTFSW}; c: I_{PTFRC})			0.175[b] 0.105[c]
PI_{Ψ} ($F \le 0.03$; $U \ge 0.06$) (I_{PTFRC} only)			0.140

F: favourable limit; U: unfavourable limit. Numbers represent the relative weight on the indicator attributed to each input.

[a] RRMSE, relative mean square error (Loague and Green, 1991):

$$\mathrm{RRMSE}=\frac{100}{\overline{M}}\left[\frac{\sum_{i=1}^{n}(E_i-M_i)^2}{n}\right]^{0.5}$$

EF, modelling efficiency index (Loague and Green, 1991):

$$\mathrm{EF}=1-\frac{\sum_{i=1}^{n}(E_i-M_i)^2}{\sum_{i=1}^{n}(M_i-\overline{M})^2}$$

r, Pearson's correlation coefficient (Addiscott and Whitmore, 1987):

$$r=\frac{\sum_{i=1}^{n}(E_i-\overline{E})(M_i-\overline{M})}{\sqrt{\sum_{i=1}^{n}(M_i-\overline{M})^2\sum_{i=1}^{n}(M_i-\overline{M})^2}}$$

PI_y, pattern index (Donatelli et al., 2004):

$$\mathrm{PI}_y=\max_{l,m=1,\ldots,3;l\neq m}\left|\frac{1}{q_l}\sum_{i_l=1}^{q_l}(E_i-M_i)_{i_l}-\frac{1}{q_m}\sum_{i_m=1}^{q_m}(E_i-M_i)_{i_m}\right|$$

where i: ith value, E: estimated value, M: measured value, $\overline{M}$: average of measured values, n: number of E_i/M_i pairs, l, m: two groups being compared, q_l, q_m: number of residuals in the groups, i_l, i_m: each value of residuals in the groups, y: D_{50} (particles' median diameter, µm), OC (organic carbon content, %), or Ψ (soil water potential, log(kPa)).

The control rules for estimating module values were based on logic relationships between inputs and outputs, expressed in linguistic terms by *if–then* statements. For example, when two input variables are aggregated, four rules are formalized as follows:

Premise	Conclusion
if x_1 is F and x_2 is F	then y_1 is B_1
if x_1 is F and x_2 is U	then y_2 is B_2
if x_1 is U and x_2 is F	then y_3 is B_3
if x_1 is U and x_2 is U	then y_4 is B_4

where x_i is an input variable, y_i is an output variable and B_i is a decision rule (or expert weight). The value of each conjunction (… *and* …) is the minimum of the quantified fuzzy sets, as obtained from complementary S-shaped distribution curves. The output fuzzy sets for all the rules are then aggregated into a single fuzzy set. This set encompasses a range of output values, and is defuzzified in order to resolve a single crisp output value from the set. The centroid method was selected to obtain the representative non-fuzzy value for the output, as commonly adopted in the Sugeno-type systems. The expert reasoning runs as follows: if all input variables are F, the value of the module is 0 (good agreement between estimates and measurements); if all indices are U, the value of the module is 1 (bad agreement), while all the other combinations assume intermediate values. In order to simplify the rules assignments, the application automatically derives the decision rules B_i from normalized weights assigned to each input variable, with a simple algorithm. These weights were chosen on the basis of the authors' own experience in handling each statistic. As a general rule, the same importance was assigned to each index within a module. The only exception was made for the module Pattern in I_{PTFRC}, where PI_{Ψ}, the pattern index computed against water potential, was given a higher importance than the other two indices (Donatelli et al., in press-b). Based on the authors' judgment, a decreasing importance was assigned to the modules: Accuracy (0.50), Pattern (0.35) and Correlation (0.15). Users are allowed to re-define fuzzy settings (limits and weights) to suit their own experience. The methodology to compute the integrated indices is fully described in Donatelli et al. (in press-b) and in IRENE_DLL's help file (code examples and applications).

3. Software application and availability

The application for PTFs evaluation was developed using mathematical routines provided by IRENE_DLL, which are invoked by macro programs written with the VBA editor of Excel. It runs over Windows 2000/XP systems, provided that the DLL is registered on the user's system. The code is extensively documented and the use of the DLL is fully illustrated in the help file distributed with the software. The Excel file, including VBA source code, is automatically installed by running IRENE_DLL's installing program (release 1.05), and freely accessible to non-profit users through the site http://www.sipeaa.it/tools or upon request to the authors. An illustrative example for using the above approach to evaluate sets of alternative PTFs is shown in Table 2 (for the indicator I_{PTFSW}). In the spreadsheet, numerical outputs are accompanied by graphs (estimates versus measurements; residuals versus independent variables), designed to ease the interpretation of the evaluation results.

4. Discussion and conclusions

The application presented is an example of a client-server architecture built re-using a fully documented component of verified quality, IRENE_DLL, to compute a number of model evaluation related statistics. The Excel client was chosen because of its wide distribution and popularity. Its flexibility has allowed the development of a simple, intuitive, and efficient user interface which can be further customized according to user's needs.

In this paper, we have outlined some of the unique computational capabilities implemented in IRENE_DLL, illustrating how its fuzzy-based methods fulfilled the requirements of a particular worked example. The strategy for the PTFs evaluation has shown excellent performance in discriminating across alternative functions. These are partial results from a subset of data, only meant to illustrate the type of outputs attainable from using PTFIndicator (concluding

Table 2
Values of basic indices (RRMSE, EF, r, $\mathrm{PI}_{D_{50}}$, $\mathrm{PI}_{\mathrm{OC}}$), modules (Accuracy, Correlation, Pattern), and integrated index (I_{PTFSW}) for evaluation of soil water estimates at field capacity and wilting point by alternative PTFs

Pedotransfer function	RRMSE	EF	r	$\mathrm{PI}_{D_{50}}$	$\mathrm{PI}_{\mathrm{OC}}$	Accuracy	Correlation	Pattern	I_{PTFSW}
Volumetric soil water content at −33 kPa (field capacity)									
Manrique	22.6	0.365	0.682	**0.011**	0.145	0.073	1.000	0.500	0.336
Rawls	18.9	0.555	**0.868**	0.028	0.065	**0.000**	**0.051**	0.500	0.177
Mayr–Jarvis	41.0	−1.092	0.686	0.013	0.137	0.635	1.000	0.500	0.651
Rawls–Brakensiek	18.7	0.566	0.855	0.026	0.075	**0.000**	0.103	0.500	0.181
HYPRES	**16.8**	**0.651**	0.851	0.030	**0.029**	**0.000**	0.122	**0.000**	**0.004**
Volumetric soil water content at −1500 kPa (wilting point)									
Manrique	**55.0**	**−0.677**	0.518	0.013	**0.032**	**0.972**	**1.000**	**0.004**	**0.649**
Rawls	78.5	−2.416	0.528	0.020	0.099	1.000	**1.000**	0.500	0.825
Mayr–Jarvis	70.0	−1.713	0.362	**0.006**	0.059	1.000	**1.000**	0.499	0.825
Rawls–Brakensiek	94.3	−3.922	0.483	0.019	0.067	1.000	**1.000**	0.500	0.825
HYPRES	73.2	−1.972	0.487	0.035	0.054	1.000	**1.000**	0.463	0.800

Values in bold are the most favourable within each group of rows (the most favourable values for the modules and the indicator are the lowest ones, see text). Calculations over subsets of the soils data from the HYPRES database (Wösten et al., 1999). Details about each function are in Donatelli et al. (in press-a).

results on the ability of the functions are in press). Even though in this study integrated indices have been only developed to evaluate PTFs of soil water content, the procedure shown can be easily adapted to develop integrated indices for evaluating PTFs of other soil variables (e.g., saturated conductivity, bulk density, etc.).

Acknowledgements

The authors gratefully acknowledge Dr. Henk Wösten of the Alterra Green World Research (Wageningen, The Netherlands) and Dr. Attila Nemés of the Research Institute for Soil Science and Agricultural Chemistry of the Hungarian Academy of Science (Budapest, Hungary) for making available the soil database used in this work.

Research was conducted under the auspices of the Italian Ministry of Agricultural and Forestry Policies, project SIPEAA, paper 24.

References

Addiscott, T.M., Whitmore, A.P., 1987. Computer simulation of changes in soil mineral nitrogen and crop nitrogen during autumn, winter and spring. Journal of Agricultural Science (Cambridge) 109, 141–157.

Bouma, J., 1989. Using soil survey for quantitative land evaluation. Advances in Soil Sciences 9, 177–213.

Bellocchi, G., Acutis, M., Fila, G., Donatelli, M., 2002. An indicator of solar radiation model performance based on a fuzzy expert system. Agronomy Journal 94, 1222–1233.

Donatelli, M., Acutis, M., Bellocchi, G., Fila, G., 2004. New indices to quantify patterns of residuals produced by model estimates. Agronomy Journal 96, 631–645.

Donatelli, M., Acutis, M., Nemés, A., Wösten, J.H.M. Integrated indices for pedotransfer function evaluation. In: Pachepsky, Y.A., Rawls, W. (Eds.), Development of Pedotransfer Functions in Soil Hydrology. Developments in Soil Science, vol. 30. Elsevier, Amsterdam, in press-a.

Donatelli, M., Wösten, J.H.M., Bellocchi, G. Evaluation of pedotransfer functions. In: Pachepsky, Y.A., Rawls, W. (Eds.), Development of Pedotransfer Functions in Soil Hydrology. Developments in Soil Science, vol. 30. Elsevier, Amsterdam, in press-b.

Donatelli, M., Acutis, M., Danuso, F., Mazzetto, F., Nasuelli, P., Nelson, R., Omicini, A., Speroni, M., Trevisan, M., Tugnoli, V., 2002. Integrated procedures for evaluating technical, environmental and economical aspects in farms: the SIPEAA project. Proceedings of the Seventh ESA, 15–18 July, Cordoba, Spain, pp. 271–272.

Fila, G., Bellocchi, G., Donatelli, M., Acutis, M., 2003. IRENE_DLL: a class library for evaluating numerical estimates. Agronomy Journal 95, 1330–1333.

Hamblin, A., 1991. Sustainable agricultural systems, what are the appropriate measures for soil structure? Australian Journal of Soil Research 29, 709–715.

He, H.S., D.R., Larsen, Mladenoff, D.J., 2002. Exploring component-based approaches in forest landscape modelling. Environmental Modelling & Software 17 (6), 519–529.

Hennings, V., Müller, U., Tietje, O., 1997. Evaluation of pedotransfer functions using the laboratory database of a soil information system. In: Bruand, A., Duval, O., Wösten, J.H.M., Lilly, A. (Eds.), The use of pedotransfer in soil hydrology research in Europe. Proceedings of the Second Workshop "Using Existing Soil Data to Derive Hydraulic Parameters for Simulation Models in Environmental Studies and in Land Use Planning", Orleans, France.

Imam, B., Sorooshian, S., Mayr, T., Schaap, M., Wösten, H., Scholes, B., 1999. Comparison of pedotransfer functions to compute water holding capacity using the van Genuchten model in inorganic soils. IGBP-DIS working Paper #22, The International Geosphere-Biosphere Programme, Stockholm, Sweden.

Loague, K., Green, R.E., 1991. Statistical and graphical methods for evaluating solute transport models: overview and application. Journal of Contaminant Hydrology 7, 51–73.

Pachepsky, Y.A., Rawls, W.J., Timlin, D.J., 1999. The current status of pedotransfer functions: their accuracy, reliability, and utility in field- and regional-scale modelling. In: Corwin, D.L., Loague, K., Ellsworth, T.R. (Eds.), Assessment of non-point source pollution in the vadose zone. Geophysical Monograph 108. American Geophysical Union, Washington D.C.

Sugeno, M., 1985. An introductory survey of fuzzy control. Information Science 36, 59–83.

Tietje, O., Hennings, V., 1996. Accuracy of the saturated hydraulic conductivity prediction by pedo-transfer functions compared to the variability within FAO textural classes. Geoderma 69, 71–84.

Tietje, O., Tapkenhinrichs, M., 1993. Evaluation of pedotransfer functions. Soil Science Society of America Journal 57, 1088–1095.

Wösten, J.H.M., 2002. Pedotransfer functions. In: Lal, R. (Ed.), The Encyclopedia of Soil Science. Dekker, New York.

Wösten, J.H.M., Lilly, A., Nemes, A., Le Bas, C., 1999. Development and use of a database of hydraulic properties of European soils. Geoderma 90, 169–185.

Wösten, J.H.M., Pachepsky, Ya.A., Rawls, W.J., 2001. Pedotransfer functions: bridging the gap between available basic soil data and missing hydraulic characteristics. Journal of Hydrology 251, 123–150.