

Short communication

IRENE: a software to evaluate model performance

Gianni Fila^{a,*}, Gianni Bellocchi^a, Marco Acutis^{b,1}, Marcello Donatelli^a

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

^b *Department of Crop Science, via Celoria 2, 20133 Milan, Italy*

Abstract

The software IRENE (Integrated Resources for Evaluating Numerical Estimates) is a data analysis tool designed to provide easy access to statistical techniques for use in model evaluation. Mostly, non-replicated model estimates (E_i) are compared against non-replicated measurements (M_i). The software also allows comparing individual estimates against replicated measurements (or vice versa) and replicated estimates against replicated measurements. The evaluation of model performance is essentially based on the difference $E_i - M_i$, or on the correlation–regression of E_i vs. M_i (or vice versa). In addition, model evaluation by probability distributions, pattern analysis, or fuzzy-based aggregation statistics is allowed. Graphics are included in most analytical tasks. The results are displayed in separate spreadsheets and can be exported into MS Excel workbooks.

© 2002 Elsevier Science B.V. All rights reserved.

Keywords: Model evaluation; Test statistics; Difference measures; Pattern analysis; Fuzzy rules

1. Introduction

The use of simulation models requires a comparison between estimated and measured data to assess model reliability (Sinclair and Seligman, 2000). Discrepancies between estimated and measured data alert users about the inadequacy of the model in addressing the scientific questions being evaluated. The analysis of simulation results is commonly based on a number of statistics (Willmott and Wicks, 1980; Fox, 1981; Addiscott and

Whitmore, 1987; Loague and Green, 1991; Mayer and Butler, 1993; Power, 1993; Smith et al., 1997; Martorana and Bellocchi, 1999; Metselaar, 1999; Kobayashi and Salam, 2000; Yang et al., 2000). Difference-based statistics quantify the departure of the model outputs from the measurements. The recently developed pattern indices (Donatelli et al., 2000) assess systematic biases of model residuals against external variables. Such indices are often used in conjunction with correlation and regression coefficients. The aggregation of more statistics was also proposed (Bellocchi et al., 2002) to capture inaccuracies of different sources into synthetic indicators.

This paper describes a dedicated software tool to be used in model evaluation, analysing model outputs by means of an integrated set of available statistical techniques.

* Corresponding author. Tel.: +39-051-6316845; fax: +39-051-374857

E-mail addresses: g.fila@isci.it (G. Fila), g.bellocchi@isci.it (G. Bellocchi), macutis@ideagate.it (M. Acutis), m.donatelli@isci.it (M. Donatelli).

¹ Tel.: +39-02-50316577; fax: +39-02-50316575.

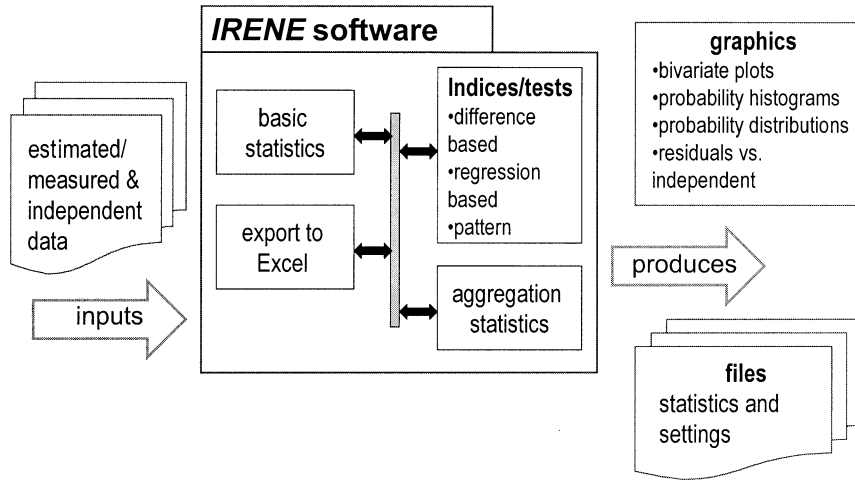


Fig. 1. Diagram representing the software IRENE.

2. Software description

IRENE (Integrated Resources for Evaluating Numerical Estimates) is a MS Windows (98/NT/2000/XP) software, that provides easy access to

model evaluation techniques. A variety of statistical tools, in a single, integrated environment, are provided for comparing model estimates (E_i) against measured data (M_i). The procedures available in IRENE and the inputs/outputs are summar-

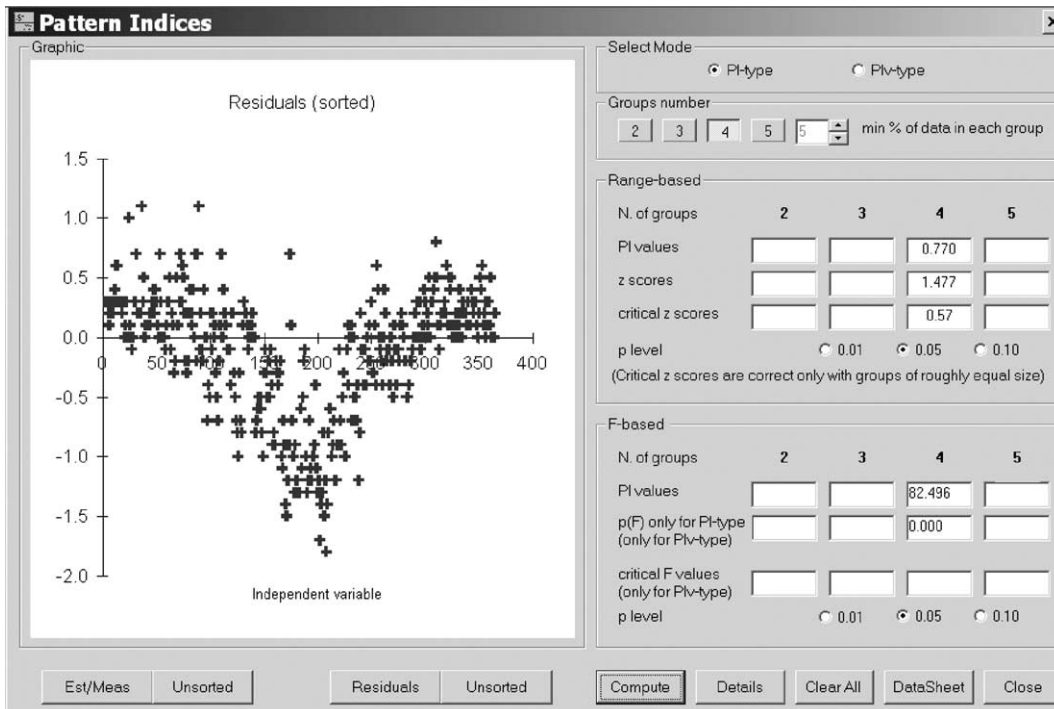


Fig. 2. Screen of pattern indices computation.

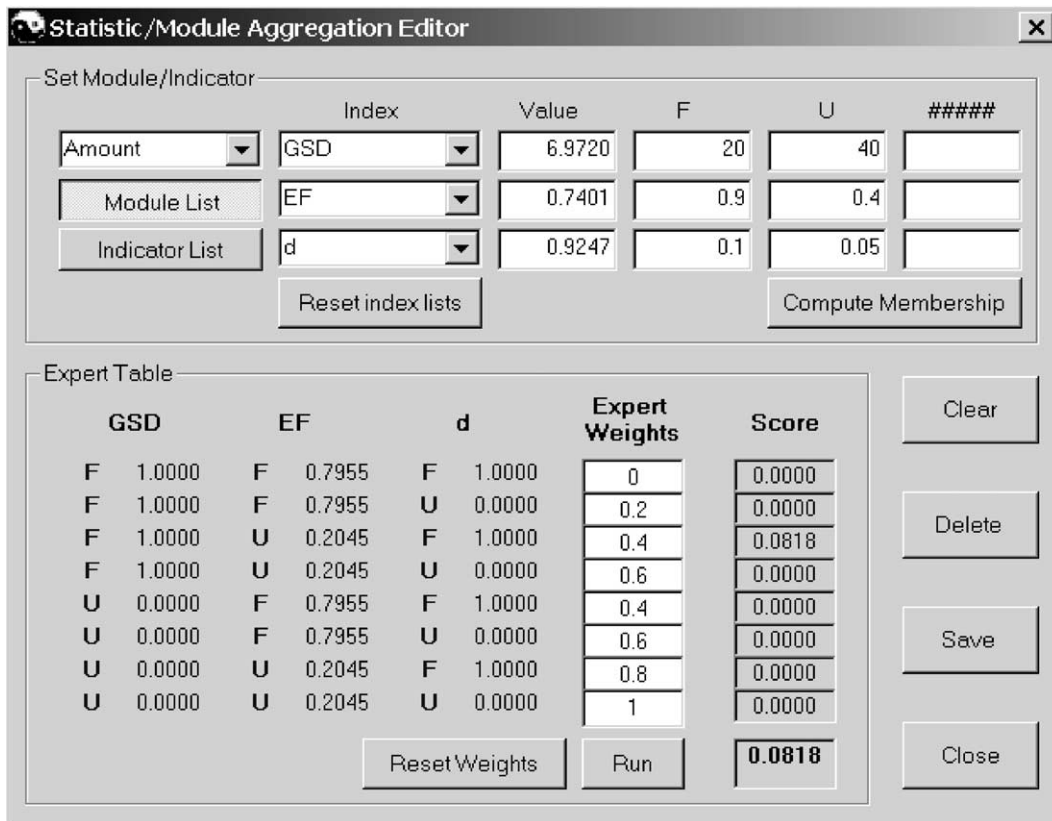


Fig. 3. Screen of fuzzy-based statistics aggregation editor.

ized in Fig. 1. A user-friendly interface allows the users to easily manipulate inputs, load selected data, calculate evaluation statistics, display graphics, and export outputs. The program reads input data from MS Excel spreadsheets, organized as columns of estimated and measured data, and independent variables. Basic statistics and graphics are available to evaluate inputs prior to analysing them. The results of the analysis can be exported into MS Excel files.

2.1. Summary of analysis capabilities

The program provides more than 40 statistics to be used in model evaluation. Modules in the software are designed to compute both statistics based on the differences $E_i - M_i$ and on the correlation–regression of E_i vs. M_i (or M_i vs. E_i). Difference-based statistics consist of simple, abso-

lute and squared statistics. Regression parameters are calculated according to two methods: ordinary least squares and reduced major axis. The latter is regarded to deal with uncertainty associated to both measurements and model estimates (Ricker, 1984).

The software allows performing four types of comparison between model outputs and measurements: non-replicated E_i against non-replicated M_i (One-to-One Evaluation), non-replicated E_i against replicated M_i (One-to-Many Evaluation), replicated E_i against non-replicated M_i (Many-to-One Evaluation), and replicated E_i against replicated M_i (Many-to-Many Evaluation). Confidence intervals are calculated and test statistics are performed when replicates are available. When both estimates and measurements are replicated, analysis of different type are executed depending on how the replicates are generated. If the pairs of

estimates and measurements are replicated, model failure/goodness is evaluated according to two variation sources: experimental error, and replication variability (Pitman–Morgan procedure).

A software component is devoted to the quantification of patterns in the residuals, by means of pattern indices (Fig. 2). Statistics of different type may be aggregated into synthetic indicators by fuzzy-based rules (Fig. 3). Cumulative probability distributions (i.e. probability of exceedence) are computed for evaluating outputs from models that include stochastic inputs.

2.2. Availability and feedback

IRENE is available free of charge for non commercial purposes. The installation package may be downloaded from: <http://www.isci.it/tools>. The program is fully documented by the user's manual, which gives detailed description of both the techniques being implemented and the scientific background. The manual is provided with the software package and is available both on-line from the IRENE interface, and as a standalone help. Comments about IRENE may be sent to agronomy@isci.it.

Acknowledgements

Version 1.00 of IRENE was developed with supports provided by the Italian Ministry of Agricultural and Forestry Policies, project SIPEAA (paper n. 5).

References

- Addiscott, T.M., Whitmore, A.P., 1987. Computer simulation of changes in soil mineral nitrogen and crop nitrogen during autumn, winter and spring. *J. Agric. Sci. (Cambridge)* 109, 141–157.
- Bellocchi, G., Acutis, M., Fila, M., Donatelli, M., 2002. An indicator of solar radiation model performance based on a fuzzy expert system. *Agron. J.* 94, 1222–1233.
- Donatelli, M., Acutis, M., Bellocchi, G., 2000. Two statistical indices to quantify patterns of errors produced by models. Proceedings of the Third International Crop Science Conference, 17–22 August, Hamburg, Germany, p. 186.
- Fox, D.G., 1981. Judging air quality model performance: a summary of the AMS workshop on dispersion models performance. *Bull. Am. Meteorol. Soc.* 62, 599–609.
- Kobayashi, K., Salam, M.U., 2000. Comparing simulated and measured values using mean squared deviation and its components. *Agron. J.* 92, 345–352.
- Loague, K., Green, R.W., 1991. Statistical and graphical methods for evaluating solute transport models: overview and application. *J. Contam. Hydrol.* 7, 51–73.
- Martorana, F., Bellocchi, G., 1999. A review of methodologies to evaluate agroecosystem simulation models. *Ital. J. Agron.* 3, 19–39.
- Mayer, D.G., Butler, D.G., 1993. Statistical validation. *Ecol. Model.* 68, 21–32.
- Metselaar, K., 1999. Auditing predictive models: a case study in crop growth. Thesis Wageningen Agricultural University, Wageningen.
- Power, M., 1993. The predictive validation of ecological and environmental methods. *Ecol. Model.* 68, 33–50.
- Ricker, W.E., 1984. Computation and uses of central trend lines. *Can. J. Zool.* 62, 1897–1905.
- Sinclair, T.R., Seligman, N., 2000. Criteria for publishing papers on crop modeling. *Field Crops Res.* 68, 165–172.
- Smith, P., Smith, J.U., Powlson, D.S., McGill, W.B., Arah, J.R.M., Chertov, O.G., Coleman, K., Franko, U., Frolking, S., Jenkinson, D.S., Jensen, L.S., Kelly, R.H., Klein-Gunnewiek, H., Komarov, A.S., Li, C., Molina, J.A.E., Mueller, T., Parton, W.J., Thornley, J.H.M., Whitmore, A.P., 1997. A comparison of the performance of nine soil organic matter models using datasets from seven long-term experiments. *Geoderma* 81, 153–225.
- Willmott, C.J., Wicks, D.E., 1980. An empirical method for the spatial interpolation of monthly precipitation within California. *Phys. Geogr.* 1, 59–73.
- Yang, J., Greenwood, D.J., Rowell, D.L., Wadsworth, G.A., Burns, I.G., 2000. Statistical methods for evaluating a crop nitrogen simulation model, N_ABLE. *Agric. Syst.* 64, 37–53.