CROP MODELLING AND VALIDATION: INTEGRATION OF IRENE_DLL IN THE WARM ENVIRONMENT

MODELLAZIONE DELLE COLTURE E VALIDAZIONE: INTEGRAZIONE DI IRENE_DLL NELL'AMBIENTE WARM

Gianni Bellocchi[1*], Roberto Confalonieri[2], Marcello Donatelli[3]

[1] *European Commission Directorate General Joint Research Centre, Institute for Health and Consumer Protection, Biotechnology and GMOs Unit, via E. Fermi 1-TP 331, I-21020 Ispra (VA), Italy*
[2] *Institute for the Protection and Security of the Citizen, Joint Research Centre of the European Commission, AGRIFISH Unit, MARS-STAT Sector, TP 268 - 21020 Ispra (VA), Italy*
[3] *Agriculture Research Council, Research Institute for Industrial Crops, via di Corticella 133, 40128 Bologna, Italy*
* *Corresponding Author : Tel.: +39 0332 786798; E-mail: gianni.bellocchi@jrc.it; fax: +39 0332 785483*

Received 28/9/2006 – Accepted 6/12/2006

Abstract

The growing importance of biophysical models in research and application-oriented projects is driving a growing interest in developing suitable approaches to evaluate model performance. Valuable validation techniques should assess the performance of complex models under a variety of conditions, and should include a wide range of validation measures. After discussing validation issues and methods currently used to assess the quality of simulation models, the integration of the software component for model output evaluation IRENE_DLL within the rice modelling system WARM is illustrated. The purpose is to demonstrate, via a case study, that great utility in validation can be gained through the implementation and use of object-oriented software tools targeting at modularity and reusability inside a modelling environment. This facilitates model validation sessions and extensibility of tools towards new approaches possibly coming out of research. These challenges can be met by using a wide range of approaches and by expanding horizons in validation whilst tailoring the evaluation requirements to the specific objectives of the model application. The availability of appropriate software tools allow actions that are not frequently executed within the context of project-based modelling activities, thus helping the dissemination of validation experiences and preventing future modelling projects from repetition of validation efforts.

Keywords: IRENE_DLL, model validation, object-oriented software, WARM

Riassunto

La crescita dell'importanza dei modelli biofisici in ambiti scientifici e applicativi sta promuovendo un crescente interesse nello sviluppo di approcci idonei a valutare le prestazioni dei modelli stessi. Tecniche appropriate devono permettere la valutazione di modelli complessi sotto diverse condizioni, attraverso una serie di indici statistici, test e grafici. Nel presente lavoro, dopo aver discusso le problematiche relative alla validazione dei modelli e dei metodi correntemente usati per valutare la qualità delle simulazioni, come caso studio è presentata la strategia di integrazione della libreria IRENE_DLL entro l'ambiente di sviluppo del modello di simulazione del riso WARM. Scopo del lavoro è illustrare le opportunità offerte da librerie per la validazione (come IRENE_DLL), sviluppate secondo la programmazione a oggetti, di tipo modulare e utilizzabili entro l'ambiente di sviluppo del modello. Questo approccio agevola l'esecuzione delle validazioni e permette di incorporare nell'ambiente di sviluppo nuovi approcci alla validazione man mano prodotti dalla ricerca. Le questioni che la validazione dei modelli pone possono essere affrontate attraverso una varietà di approcci ed espandendo gli orizzonti della validazione ridefinendo le richieste di bontà del modello rispetto a specifiche applicazioni. La disponibilità di strumenti software adeguati permette di intraprendere azioni comunemente non eseguite nel contesto della comune attività di modellazione, facilitando anche la disseminazione delle esperienze di validazione e prevenendo la reiterazione delle stesse validazioni in progetti successivi.

Parole chiave: *IRENE_DLL, validazione dei modelli, software orientato agli oggetti, WARM*

Introduction

The evaluation of model adequacy is an essential step of the modelling process because it indicates the level of accuracy of the model estimations (how closely model-estimated values are to the actual values). This is an important phase either to build up confidence on the current model or to allow selection of alternative models out of a list of potential candidates. The concept of model validation, in spite of controversial terminology (Konikow and Bredehoeft, 1992; Bredehoeft and Konikow, 1993; Bair, 1994; Oreskes, 1998), is quite generally accepted and interpreted in terms of assessment of model suitability for a particular purpose, that means a model is valid and sound if it accomplishes what is expected of it (Forrester, 1961; Hamilton, 1991; Landry and Oral, 1993; Rykiel Jr., 1996; Sargent, 2001). One of the principles of validating models dictates that complete testing is not

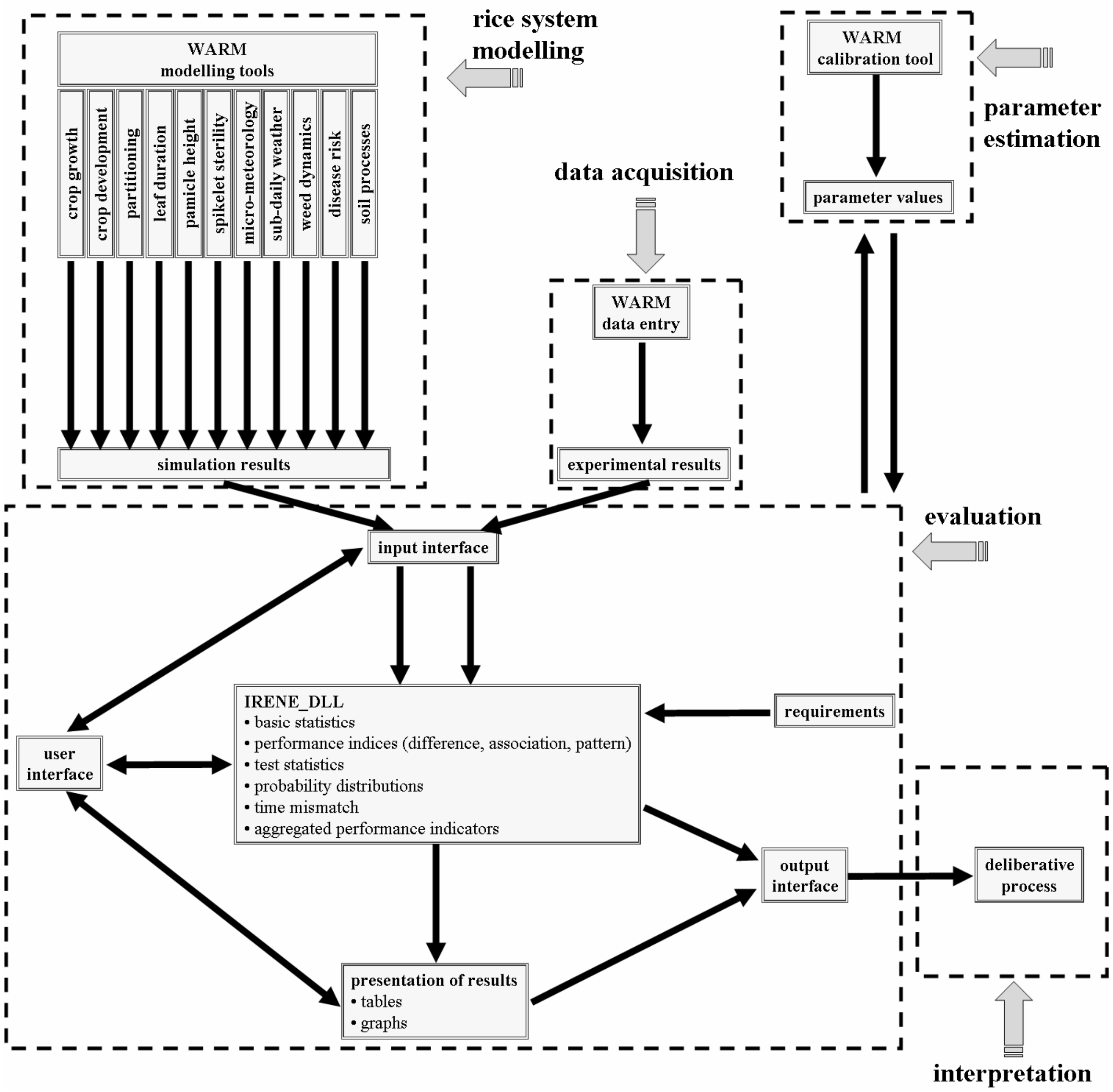

Fig. 1 - Integration of IRENE_DLL within WARM environment.

Fig. 1 - *Integrazione di IRENE-DLL entro l'ambiente di WARM.*

possible (Balci, 1997), thus to prove that a model is absolutely valid is an issue without a solution. Exhaustive validation requires testing all possible model outputs under all virtually possible inputs (i.e. conditions). In practice, model validation aims at increasing confidence in model accuracy as much as possible, which is partially determined by the intended uses of a specific model and project objectives. Model evaluation can be performed with different levels of detail, but there is no formally described procedure that sets quantitatively minimum requirements for a model testing procedure. The low priority often assigned to validation in model project proposals and development plans indicates a tendency towards the minimum standard to be adopted. As matter of fact, in spite of the several validation methods available, only a limited number of methods are commonly used in modelling projects, often due to time and resource constraints. This is also because different users may have different thresholds for confidence. In general, a limited testing may hinder modeller's ability to substantiate model accuracy sufficiently.

In this paper model validation methods and techniques are reviewed and discussed in general and in their effectiveness to support the modelling project WARM (Confalonieri *et al.*, 2005).

Numerical indices and test statistics

Evaluation of simulated versus observed outputs by means of numerical indices and test statistics is an accepted action of the modelling practice. Mean bias (MB), the mean difference between model estimates and observations, is likely to be the oldest statistic to assess model accuracy (Cochran and Cox, 1957). MB is quite used in model validation, but one statistic that normally takes precedence over the others is the mean square error (MSE), or equivalently its square root, the root mean square error (RMSE, or derived statistics such as the relative root mean square error RRMSE). This is also the

Tab. 1 - Summary list of functions implemented in IRENE_DLL.
Tab. 1 - *Schema riassuntivo delle funzioni implementate in IRENE_DLL.*

Difference-based analysis	Statistical association-based analysis	Pattern analysis	Probability distributions	Statistics aggregation	Time mismatch
Simple differences	Least squares method	Range-based pattern indices	Probability density functions	First-level aggregation	Time mismatch indices
Square differences	Reduced major axis method	F-based pattern indices	Cumulative distribution functions	Second-level aggregation	
Absolute differences					
Test statistics					

statistic whose value is usually minimized during the parameter calibration process (e.g. Sorooshian *et al.*, 1983, 1993; Wallach, 1999). Mean absolute error (MAE) measures the mean absolute difference between model estimates and observations (Mayer and Butler, 1993), and it is also used as the mean absolute percent error. The modelling efficiency statistic (EF), initially used in hydrology models (e.g. Loague and Green, 1991), is interpreted as the proportion of variation explained by the model.
A linear regression between estimated and observed values (or vice versa) is also commonly used. The hypothesis is that the regression passes through the origin and has a slope of unity (Dent and Blackie, 1979). Nonetheless, the use of the least-square method to derive a linear regression of observed on modelled values for model evaluation has little interest since the estimated value is useless in evaluating the mathematical model; therefore the r^2 (goodness-of-fit) is irrelevant since one does not intend to make predictions from the fitted line (Mitchell, 1997). Additionally, necessary assumptions (normality, homoscedasticity, independence, X-axis values known without errors) have to be considered when performing a linear regression. The r coefficient and its counterpart r^2 require careful interpretation because high values of such coefficients in isolation do not indicate model accuracy and do not imply that the estimated regression line is a good fit of the model estimation. The studentized statistics of the intercept and slope is t-distributed, and can be used to check for intercept=0 and slope=1, but there are several concerns about the appropriateness of linear regression in model evaluation (Harrison, 1990; Mitchell, 1997).
The comparison of the distribution of the observed and estimated values has also been utilized to identify model adequacy for stochastic (Reynolds and Deaton, 1982) and deterministic models (Dent and Blackie, 1979). The common Kolmogorov-Smirnov's D test has been used to assess the probability that two data sets (observations and simulations) have the same distribution. It consists to measure the overall difference of the area between two cumulative distribution functions (Press *et al.*, 1992).
Several other criteria for comparing models have been proposed and discussed to appraise goodness-of-fit of simulation outcome (e.g. Bellocchi, 2004; Krause *et al.*, 2005; Tedeschi, 2006). No single statistic, although highlighting particular aspects of the comparison, is in general sufficient to adequately assess the performance of the models when comparing observed and estimated values. Therefore, the performance of models can be thoroughly assessed if a group of criteria is established prior the comparison of the models. Under single-metric validation capabilities (non-integrated assessment), users may be unable to claim sufficient accuracy of a large and complex model application due to model complexity, reliance on qualitative human judgment, and lack of complete testing. Many authors believe that there is no robust statistic or graphical representation which can be used to draw conclusions in model evaluation and therefore several methods need to be used together to give a comprehensive check (e.g., Yang *et al.*, 2000). The use of multiple metrics might be helpful in this respect. Since most of the statistics used are based on the difference or the association between calculated and measured data, as emphasized by Donatelli *et al.* (2002b) there is little attention by many users for aspects related to the patterns of residuals and time series comparison. Some examples of integrated, multi-metric validation including pattern analysis and time series investigations actually exist: Bellocchi *et al.* (2002, 2003, 2004), Donatelli *et al.* (2002a, 2004a, b), Rivington *et al.* (2005), Diodato and Bellocchi (2007).

Computer-aid to model validation

Complex biophysical models implement rate equations, comprise approaches with different levels of empiricism, make use of parameters partially auto-correlated, aim at simulating systems which show a non-linear behavior, and often require numerical rather than analytical solutions. Therefore, the computer program, including technical issues and possible errors, is tested rather than the mathematical model representing the system (Leffelaar *et al.*, 2003). Each version of a model, throughout its development life cycle, should be subjected to output testing, designed by identifying test scenarios, test cases, and/or test data. Applying the same test to each model release is repetitive and time consuming, requiring the preservation of the test scenarios, test cases, and test data for re-use.
Software tools, specifically created for model validation, provide an effective support because can significantly reduce the testing time and effort. Software tools collecting numerical indices and test statistics for model validation do actually exist, e.g. IRENE (Integrated Resources

for Evaluating Numerical Estimates, Fila *et al.*, 2003a), Model Validation Kit (Olesen and Chang, 2005), ModEval (Tedeschi, 2006).
To meet the substantial model quality challenges, it is however necessary to improve the current tools, technologies, and their cost-benefit characterizations. The emergence of new technologies in simulation modelling has in fact fostered debate on the re-use of modelling tools. In order to possibly include legacy data sources into newly developed systems, object-oriented development has emerged steadily as a paradigm that focuses on granularity, productivity and low maintenance (Timothy, 1997). There is some consensus (e.g. Glasow and Pace, 1999) that object-oriented development is indeed an effective and affordable way of creating model applications and conducting model validation. Particular emphasis should be placed on designing and coding object-oriented simulation models to properly transfer simulation control between entities, resources and system controllers and on techniques for obtaining an easily verifiable correspondence between simulation code and system behavior. It is crucial, therefore, to consider the issue of validity when considering model re-use as it needs to be a fundamental part of any re-use strategy. The distribution of validated model objects can substantially decrease the model validation effort when re-used within different modelling environments. A key step in this direction is the coupling between model objects (whole model and sub-models) and validation techniques (the latter also implemented into object-oriented software), that allows objects to communicate between each other. The validation system should stand at the core of a structure where both the modelling system and a data provider supply inputs to the validation system. The freeware, Microsoft COM-based tool IRENE_DLL (Integrated Resources for Evaluating Numerical Estimates_Dynamic Link Library, Fila *et al.*, 2003b; available for downloading through the web site http://www.sipeaa.it/tools), is a flexible tool providing extensive, integrated, statistical capabilities (Table 1). The modular structure allows it to be plugged into existing model application software. Validation runs can be automated and executed at any time objects are added or modified, using a wide range of integrated statistics. The DLL was used to tailor a dedicated application for evaluation of pedotransfer functions (Fila *et al.*, 2006). However, IRENE_DLL's capabilities are still far from being fully exploited. Since IRENE_DLL's was developed, the COM paradigm (http://www.microsoft.com/com) became surmounted by the .NET platform of Windows (http://www.microsoft.com/net), at the same time offering the possibility for inter-compatibility across operating systems (i.e.Mono, http://www.mono-project.com). A re-design of IRENE_DLL is under development (project SEAMLESS, http://www.seamless-ip.org, Donatelli M., CRA-ISCI, Bologna, Italy, personal communication) to provide third parties with the capability of extending methodologies without re-compiling the component. This will ensure greater transparency and ease of maintenance, also providing functionalities such as the test of input data versus their definition prior to computing any simple or integrated validation metric.

The coupling WARM-IRENE_DLL

WARM (Water Accounting Rice Model) is a modelling environment for simulation of flooded rice, managing the most relevant issues of crop development, growth and production on a daily time step (Confalonieri *et al.*, 2005). The ready-to-use software environment of WARM includes supporting tools for sensitivity/uncertainty analysis, manual/automatic parameter calibration, and model validation. The latter tools in particular are provided by IRENE_DLL, which also interact with parameter calibration tools by supplying the calibrator with metrics as cost functions to be optimized. Figure 1 illustrates the coupling strategy between crop modelling/data acquisition tools and IRENE_DLL's validation methods. The DLL communicates with both the modelling part and the data provider via a suitable protocol and allow the user to interact (via a dedicated graphical user interface). The output coming out of the validation system can be offered to a deliberative process for interpretation of results that engage model developers and users.
Adjustments in the modelling system or critical reviewing of data used to validate the model can be made afterwards, if the results are assessed as not satisfactory for the purpose of modelling. A new validation-interpretation cycle can be run any time new versions of the modelling system are developed and plugged to the validation component.

Remarks

The scope and capabilities of validation techniques and tools have improved with time. However, it can be argued that the rate of evolution of validation techniques has not been as rapid as the increase in modelling capabilities. Whilst a range of modelling approaches do actually exist for improving model estimates, the fundamental problem of how numerical values produced by models can be best evaluated remains an issue. Greater value can be gained through the combined use of multiple statistical indices, metrics and graphical representations to achieve a robust form of testing. Advancements in these numerical testing methodologies for validation need to be put into structured frameworks comprised of processes such as sensitivity and uncertainty analyses (parameter and input appraisal), parameter optimization, model structure evaluation (expert review), software testing, etc. As such, validation must be seen as an integral part of the overall model development and application process, and techniques to validate models need to be developed at the same pace with which the models themselves are created, improved and applied.
WARM is an example of disciplined approach where an assorted set of validation metrics (as supplied by IRENE_DLL) are integrated within the modelling environment, taking advantage of the modular, object-oriented programming features on both sides of model and evaluation tools. To the best of the authors' knowledge, WARM-IRENE_DLL coupling is a unique example of crop modelling-validation software integration and, at the same time, a reference point of remark for future development of modelling environments. The point is the need to incorporate automated validation checks

into modeling environments using embedded software tools, with the aim of achieving greater cost and time efficiency and higher level of model credibility. The return on such an investment can easily be realized precluding the failures of modelling projects by preventing wrong simulation-based decisions.

References

Bair, E.S., 1994. *Model (in)validation - a view from courtroom. Ground Water 32, 530-531.*

Balci, O., 1997. Principles of simulation model validation, verification, and testing. Transactions of the Society for Computer Simulation International 14, 3-12.

Bellocchi, G., 2004. Appendix A: Numerical indices and test statistics for model evaluation. p. 394-400. In: Pachepsky, Y., Rawls, W. (eds.), Development of pedotransfer functions in soil hydrology. Elsevier, Amsterdam, The Netherlands.

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.

Bellocchi, G., Donatelli, M., Fila, G., 2003. Calculating reference evapotranspiration and crop biomass using estimated radiation inputs. Italian Journal of Agronomy 7, 95-102.

Bellocchi, G., Smith, J., Donatelli, M., Smith, P., 2004. Improvements in time mismatch analysis of model estimates. In: Jacobsen, S.E., Jensen C.R., Porter J.R. (Eds.), Proceedings of 8th European Society for Agronomy Congress, 11-15 July, Copenhagen, Denmark, 221-222.

Bredehoeft, J.D., Konikow, L.F., 1993. Ground water models: validate or invalidate. Ground Water 3, 178-179.

Cochran, W.G., Cox, G.M., 1957. Experimental design. John Wiley & Sons, New York, NY, USA.

Confalonieri R., Acutis M., Donatelli M., Bellocchi G., Mariani L., Boschetti M., Stroppiana D., Bocchi S., Vidotto F., Sacco D., Grignani C., Ferrero A., Genovese G., 2005. WARM: a scientific group on rice modelling. Italian Journal of Agrometeorology 2, 54-60.

Dent, J.B., Blackie, M.J., 1979. Systems simulation in agriculture. Applied Science, London, United Kingdom.

Diodato, N., Bellocchi, G., 2007. Modelling solar radiation over complex terrains using monthly climatological data. Agricultural and Forest Meteorology (in press.)

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

Donatelli, M., Acutis, M., Fila, G., Bellocchi, G., 2002a. A method to quantify time mismatch of model estimates. In: Villalobos, F.J., Testi, L. (Eds.), Proceedings of 7th European Society for Agronomy Congress, 15-18 July, Cordoba, Spain, 269-270.

Donatelli, M., Acutis, M., Nemes, A., Wösten, H., 2004b. Integrated indices for pedotransfer function evaluation. In: Pachepsky, Ya., Rawls, W.J. (Eds.), Development of pedotransfer functions in soil hydrology. Elsevier, Amsterdam, The Netherlands, 363-390.

Donatelli, M., van Ittersum, M.K., Bindi, M., Porter, J.R., 2002b. Modelling cropping systems - highlights of the symposium and preface to the special issues. Special issue I, European Journal of Agronomy 18, 1-11.

Fila G., Bellocchi, G., Acutis, M., Donatelli, M., 2003a. IRENE: a software to evaluate model performance. European Journal of Agronomy 18, 369-372.

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

Fila, G., Donatelli, M., Bellocchi, G., 2006. PTFIndicator: an IRENE_DLL-based application to evaluate estimates from pedotransfer functions by integrated indices. Environmental Modelling and Software 21, 107-110.

Forrester, J.W., 1961. Industrial dynamics. MIT Press, Cambridge, MA, USA.

Glasow, P.A., Pace, D.K., 1999. SIMVAL '99: making VV&A effective and affordable workshop. The Simulation Validation Workshop 1999, January 26-29, Laurel, MD, USA.

Hamilton, M.A., 1991. Model validation: an annotated bibliography. Communications in Statistical Theory and Methodology 20, 2207-2266.

Harrison, S. R., 1990. Regression of a model on real-system output: an invalid test of model validity. Agricultural Systems 34, 183-190.

Konikow, L.F., Bredehoeft, J.D., 1992. Ground water models cannot be validated. Advances in Water Resources 15, 75-83.

Krause, P., Boyle, D.P., Bäse, F., 2005. Comparison of different efficiency criteria for hydrological model assessment. Advances in Geosciences, 5, 89-97.

Landry, M., Oral, M., 1993. In search of a valid view of model validation for operations research. European Journal of Operational Research 66,161-167.

Leffelaar, P.A., Meike, H., Smith, P., Wallach, D., 2003. Modelling cropping systems – highlights of the symposium and preface to the special issues. 3. Session B. Model parameterisation and testing. European Journal of Agronomy 18, 189-191.

Loague, K., Green, R.E., 1991. Statistical and graphical methods for evaluating solute transport models: overview and application. J. Contam. Hydrol. 7, 51-73.

Mayer, D.G., Butler, D.G., 1993. Statistical validation. Ecological Modelling 68, 21-32.

Olesen, H.R., Chang, J.C., 2005. Consolidating tools for model evaluation. In: Skouloudis, A.N., Kassomenos, P., Bartzis, J. (Eds.), Proceeding of the 10th International Conference on Harmonisation within Atmospheric Dispersion Modelling for Regulatory Purposes, 17-20 October, Sissi, Greece, 263-267.

Oreskes, N., 1998. Evaluation (not validation) of quantitative models. Environmental Health Perspectives 106, 1453-1460.

Mitchell, P.L., 1997. Misuse of regression for empirical validation of models. Agricultural Systems 54, 313-326.

Press, W.H., Teukolsky, S.A., Vetterling, W.T., Flannery, B.P., 1992. Numerical recipes in Fortran 77: The art of scientific computing (2nd ed.). Cambridge University Press, United Kingdom.

Reynolds, J.M.R., Deaton, M.L., 1982. Comparison of some tests for validation of stochastic simulation models. Communications in Statistical Simulation and Computation 11, 769-799.

Rivington, M., Bellocchi, G., Matthews, K.B., Buchan, K., 2005. Evaluation of three model estimations of solar radiation at 24 UK stations. Agricultural and Forest Meteorology 135, 228-243.

Rykiel Jr., E.J., 1996. Testing ecological models: the meaning of validation. Ecological Modelling 90, 229-244.

Sargent, R.G., 2001. Verification, validation and accreditation of simulation models. In: Peters, B.A., Smith, J.S., Medeiros, D.J., Rohrer, M.W. (Eds.), Proceedings of the 2001 Winter Simulation Conference, December 10-13, Arlington, VA, USA, 106-114.

Soroshiaan, S., Duan, Q., Gupta, V.K., 1993. Calibration of rainfall-runoff models: application of global optimization to Sacramento Soil Moisture Model. Water Resources Research 29, 1185-1194.

Sorooshian, S., Gupta, V.K., Fulton, J.L., 1983. Evaluation of maximum likelihood parameters estimation techniques for conceptual runoff models: influence of calibration data variability and length on model credibility. Water Resources Research 1, 251-259.

Tedeschi, L.O., 2006. Assessment of the adequacy of mathematical models. Agricultural Systems 89, 225-247.

Timothy, B., 1997. An introduction to object-oriented programming. 2nd edition. Addison-Wesley, Reading, MS, USA.

Wallach, D., 1999. Linking model validation, uncertainty analysis and parameter adjustment. In: Donatelli, M., Stöckle, C.O., Villalobos, F., Villar J.M. (Eds.), Proceedings of the 1st International Symposium on Modelling Cropping Systems, June 21-23, Lleida, Spain, 271-272.

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. Agricultural Systems 64, 37-53.