

CORRELATION BETWEEN EVALUATION MODELS INDICATORS

CORRELAZIONE TRA GLI INDICATORI DI VALUTAZIONE DEI MODELLI

Alvaro Rocca^{1*}, Gianni Bellocchi², Andrea Giussani¹, Mattia Sanna¹, Alessia Perego¹, Mattia Fumagalli¹, Marco Carozzi¹, Marcello Chiodini¹, Simone Bregaglio¹, Roberto Confalonieri¹, Marco Acutis¹

¹ DISAA – Università degli studi di Milano, via Celoria 2, 20133, Milano

² INRA UREP, 5 chemin de Beaulieu, 3039, Clermont-Ferrand France, Cédex 2

*alvaro.rocca@unimi.it

Abstract

From the seventies on, many indices have been proposed for quantifying the agreement between observed and simulated data, together with new indicators for assessing model complexity, robustness and plasticity. Model evaluation is an key step in the agro-meteorology modelling process and it is one of the issues which has mostly interested the modellers community in the last years. This study tries to better understand the interrelationship between different evaluation indices through the application of correlation analysis techniques. Thirty indices of model performance were calculated for different pairs of simulated and observed data for daily solar radiation, hourly air relative humidity and crop production models. The final aim of this research is the identification of the most useful set of indicators for assessing model performance.

Keywords: evaluation, model, correlation, indicator, performance.

Parole chiave: valutazione, modelli, correlazione, indicatori, performance.

Introduzione

La crescente applicazione di modelli è alla base di studi di impatto su larga scala come dei cambiamenti climatici e comporta la necessità di gestire grandi volumi di dati simulati. Per garantire la qualità della simulazione è di fondamentale importanza la corretta valutazione dei modelli. Ci sono almeno due limiti nella ricerca esistente sulla valutazione dei risultati della simulazione. La prima, riguarda la mancanza di una metodologia accettata di opportune strategie di validazione, necessaria a garantire la riproducibilità e la validità generale dei risultati. In secondo luogo, non vi è alcuna misura accettata di come valutare le prestazioni dei modelli. Dagli anni Settanta sono stati proposti una serie di indicatori per quantificare la vicinanza tra dati osservati e simulati (Wallach, 2006; Moriasi *et al.*, 2007). Alcuni approcci si basano sulle differenze al quadrato o assolute come misure di performance, altri optano per la regressione o utilizzano indicatori di efficienza di modellazione. In generale, le misure di performance sono suddivisi in tre gruppi: accuratezza, complessità e robustezza. La valutazione del modello, per ciò riveste un passo fondamentale nel processo di modellazione (Huth e Holzworth, 2005) ed è uno dei temi che ha maggiormente interessato la comunità dei modellisti negli ultimi anni (Bellocchi *et al.*, 2010; Alexandrov *et al.*, 2011; Richter *et al.*, 2013). Questo studio mira a far luce sulle interrelazioni tra le diverse metriche di valutazione, attraverso l'applicazione di tecniche di analisi di correlazione. Sono stati calcolati trenta indicatori di performance su diverse coppie di dati simulati e osservati. Lo scopo finale è identificare la serie più utile di indicatori per valutare le prestazioni del modello.

Materiali e Metodi

Per la determinazione della correlazione sono stati considerati gli indicatori di valutazione presenti in Tab. 1. Questi indicatori includono differenze semplici, assolute e al quadrato. Inoltre, sono stati calcolati anche indici non parametrici. Si è voluto confrontare mediante correlazione gli indicatori calcolati per:

– tre modelli di radiazione solare (Hargreaves, Bristow-Campbell e Campbell-Donatelli) per quattro località (Minoprio, Ostriglia Palidano, S. Angelo Lodi e Treviglio) per tre anni (2005-2008).

Tab. 1 - Evaluation indicators considered.

Tab. 1 - Indicatori di valutazione considerati.

Evaluation index	
Mean bias error - (MBE)	Addiscott and Whitmore, 1987
Percent bias - (PBIAS)	Gupta et al., 1999
Relative error - (E)	Addiscott and Whitmore, 1987
Coefficient of residual mass - (CRM)	Loague and Green, 1991
Fractional bias - (FB)	Kumar, 2000
Maximum error - (MaxE)	Schaeffer, 1980
Maximum percent error (MaxE%)	Schaeffer, 1980
Maximum absolute error - (ME)	Loague and Green, 1991
Mean absolute error - (MAE)	Schaeffer, 1980
Relative mean absolute error (RMAE)	Mayer and Butler, 1993
Maximum absolute percent error - (MA%E)	Schaeffer, 1980
General absolute standard deviation - (GASD)	Jørgensen et al., 1991
Modified modelling efficiency - (EF ₁)	Yang et al., 2000
Simulation bias - (SB)	Kobayashi and Salam, 2000
Mean square error - (MSE)	Makowski et al., 2006
Root mean squared error - (RMSE)	Fox, 1981
Root mean squared variation - (RMSV)	Kobayashi and Salam, 2000
General standard deviation - (GSD)	Jørgensen et al., 1991
Relative root mean squared error - (RRMSE)	Robertson et al., 2002
Normalized mean squared error - (NMSE)	Kumar, 2000
Modelling efficiency - (EF)	Nash and Sutcliffe, 1970
Modelling percent efficiency - (EF%)	Greenwood et al., 1985
Persistence model efficiency - (PME)	Gupta et al., 1999
Pearson's correlation coefficient - (r)	Addiscott and Whitmore, 1987
Coefficient of determination - (CD)	Loague and Green, 1991
Index of agreement - (d)	Willmott and Wicks, 1980
Observations standard deviation ratio - (RSR)	Chu and Shirmohammadi, 2004
Median absolute error - (MdAE)	Zacharias et al., 1996
Relative median absolute error - (RMdAE)	Zacharias et al., 1996
Robust modelling efficiency - (REF)	Zacharias et al., 1996

I dati osservati provengono dalle stazioni meteorologiche ERSAF - Ente Regionale per i Servizi all'Agricoltura e alle Foreste, – otto modelli per la stima della umidità relativa dell'aria a passo orario come riportato da Bregaglio *et al.* (2010).

Con i trenta indicatori calcolati per i 28 casi è stata calcolata una matrice di correlazione e successivamente fatta una cluster analisi di tipo gerarchico utilizzando il metodo del legame medio tra i gruppi e adoperando la correlazione come misura della distanza.

Risultati

Il dendrogramma (Fig. 1) ottenuto dalla cluster analisi evidenzia come ci siano delle forte aggregazioni a diversi livelli. Da una prima analisi si osserva un'aggregazione in due gruppi già dalla decima classe di distanza. Esaminando più nel dettaglio, prima

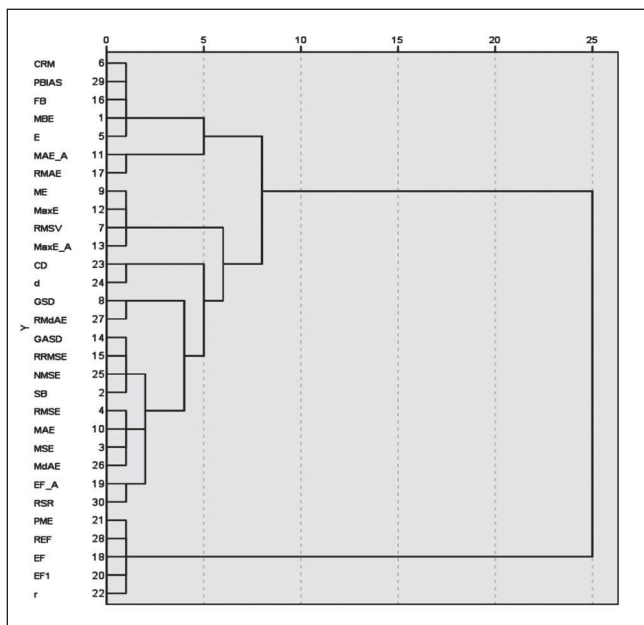


Fig. 1 - Cluster analisi dendrogram.
Fig. 1 - Dendrogramma della cluster analisi.

del quinto livello di distanza, si osservano, come gli indicatori di efficienza del modello (EF, EF₁, PME, ecc.) appartengono allo stesso gruppo. Inoltre è possibile evidenziare come gli indicatori di bias e gli indicatori basati sulla regressione sono raggruppati indipendentemente.

Conclusioni

Alla luce dei risultati ottenuti si ritiene necessario incrementare la casistica della tipologia dei dati utilizzati e considerando anche altri tipi di informazioni (ad esempio: umidità del suolo, produzione della coltura). Comunque, si ritiene che lo approccio possa essere utile anche in un'ottica di aggregazione di più indicatori in un unico indice sintetico. Infatti non vi sono prove che suggeriscono che i modelli con un buon valore per un indicatore siano altrettanto performanti per altri. Ricerche precedenti suggerivano l'utilizzo del giudizio di esperti per esplorare l'importanza delle metriche di valutazione (Rivington *et al.*, 2005), tuttavia questo approccio potrebbe essere perfezionato utilizzando la metodologia descritte in questo studio.

Bibliografia

Addiscott, T.M. and A.P. Whitmore. 1987. Computer simulation of changes in soil mineral nitrogen and crop nitrogen during autumn, winter and spring. *J. Agric. Sci. (Cambr.)*, 109: 141-157.

Alexandrov G.A, Ames D., Bellocchi G., Bruen M., Crout N., Erechtkoukova M., Hildebrandt A., Hoffman F., Jackisch C., Khaite P., Mannina G., Matsunaga T., Purucker S.T., Rivington M., Samaniego L., 2001. Technical assessment and evaluation of environ. models and software. *Environ. Modell. Softw.*, 26, 328-336.

Bregaglio S., Donatelli M., Confalonieri R., Acutis M. and Orlandini S., 2010. An integrated evaluation of thirteen modelling solutions for the generation of hourly values of air relative humidity. *Theor. Appl. Climatol* 102:429-438.

Chung, S. W.; Gassman, P. W.; Kramer, L. A.; Williams, J. R.; Gu, R. 1999. Validation of EPIC for two watersheds in Southwest Iowa. *J. Environ. Qual.* 28:971-979.

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.

Greenwood, D.J., J.J. Neeteson, and A. Draycott. 1985. Response of potatoes to N fertilizer: dynamic model. *Plant Soil*, 85: 185-203.

Gupta, H. V., S. Sorooshian, and P. O. Yapo. 1999. Status of automatic calibration for hydrologic models: Comparison with multilevel expert calibration. *J. Hydrologic Eng.* 4(2) :135-143.

Gupta, H. V., S. Sorooshian, and P. O. Yapo. 1999. Status of automatic calibration for hydrologic models: Comparison with multilevel expert calibration. *J. Hydrologic Eng.* 4(2): 135-143.

Huth N., Holzworth D. 2005 Common sense in model testing, in: Zenger A., Argent R.M. (Eds.), *Proc. MODSIM 2005 International Congress on Modelling and Simulation: Advances and applications for management and decision making*, 12–15 Dec., Melbourne, Australia, 2804-2809.

Jørgensen, S.E., L. Kamp-Nielsen, T. Christensen, J. Windolf-Nielsen, and B. Westergaard. 1986. Validation of a prognosis based upon a eutrophication model. *Ecol. Model.*, 35:165-182.

Kobayashi, K., and M.U. Salam. 2000. Comparing simulated and measured values using mean squared deviation and its components. *Agron. J.*, 92:345-352.

Kumar, A. 2000. Dispersion and risk modeling. Department of Civil Engineering, University of Toledo.

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

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

Moriasi D. N., Arnold J. G., Van Liew M. W., Bingner R. L., Harmel R. D., Veith T. L., 2007. Model evaluation guidelines for systematic quantification of accuracy in watershed simulations. *American Society of Agri. and Biological Engineers*. Vol. 50(3) : 885-900.

Nash, J. E. and J. V. Sutcliffe. 1970. River flow forecasting through conceptual models part I - A discussion of principles. *Journal of Hydrology*, 10 (3), 282-290.

Richter, K., Atzberger, C., Hank T.B., and W. Mauser. 2012. Derivation of biophysical variables from Earth observation data: validation and statistical measures. *J. Appl. Remote Sens.* 6: 063 557-2.

Rivington M., Bellocchi G., Matthews K.B., Buchan K., 2005. Evaluation of three model estimations of solar radiation at 24 UK stations. *Agr. and Forest Meteorology* 132 228-243.

Schaeffer, D.L. 1980. A model evaluation methodology applicable to environmental assessment models. *Ecol. Model.*, 8:275-295.

Wallach, 2006. Evaluating crop models. In: Wallach D., Makowski D. and Jones J.W., 2006. Working with dynamic crop models. Elsevir, Oxford.

Willmott, C.J., and D.E. Wicks. 1980. An empirical method for the spatial interpolation of monthly precipitation within California. *Phys. Geogr.*, 1:59-73.

Yang, J., D.J. Greenwood, D.L. Rowell, G.A. Wadsworth, and I.G. Burns. 2000. Statistical methods for evaluating a crop nitrogen simulation model, N_ABLE. *Agric. Syst.*, 64:37-53.

Zacharias, S., C.D. Heatwole, and C.W. Coakley. 1996. Robust quantitative pesticide transport models. *Trans. ASAE*, 39: 47-54.