

# VALUTAZIONE INTEGRATA DI MODELLI DI SIMULAZIONE: PROPOSTA E TEST DI UN INDICE DI ROBUSTEZZA

R. Confalonieri<sup>1</sup>, S. Bregaglio<sup>1</sup>

<sup>1</sup>Università degli Studi di Milano, Dipartimento di Produzione Vegetale, [roberto.confalonieri@unimi.it](mailto:roberto.confalonieri@unimi.it)

## Abstract

La valutazione dei modelli di simulazione è fondamentale per poter stilare una classifica dei modelli stessi sulla base di uno o più criteri, scelti di volta in volta a seconda degli obiettivi da raggiungere e delle condizioni di applicazione. La possibilità di valutare i modelli in modo efficiente e mirato permette di scegliere il modello più adatto per un determinato obiettivo sulla base, ad esempio, del livello di scala considerato e della disponibilità di dati. I metodi tradizionali per la valutazione dei modelli, per lo più basati su indici per la quantificazione dell'errore commesso dal modello nel riprodurre i dati osservati, sono utili solo per valutare l'accuratezza di un modello. In questo lavoro, viene presentato per la prima volta un indice di robustezza dei modelli di simulazione, basato sul rapporto tra la variabilità dell'efficienza di modellizzazione e quella delle condizioni nelle quali il modello viene utilizzato. L'indice proposto viene (i) testato su diversi modelli di tipo agrometeorologico e (ii) valutato in termini di stabilità entro anno ed entro località.

## Introduzione

La modellistica di simulazione è uno strumento fondamentale per l'analisi e la gestione dei sistemi colturali, sia per gli aspetti produttivi che per quelli legati all'impatto ambientale. La valutazione dei modelli sulla base di criteri oggettivi consente di stilare classifiche che permettano di selezionare il modello più adeguato per un determinato obiettivo. Negli ultimi anni si è assistito ad una notevole evoluzione nelle tecniche di valutazione di modelli, con la nascita di numerosi indici per la quantificazione di diverse tipologie di errore, di criteri per l'aggregazione di diversi indici e di software dedicati (Fila et al., 2003). Confalonieri et al. (2009) hanno recentemente proposto un criterio per la valutazione integrata di modelli, basato su indici di accuratezza, di complessità e di rilevanza.

L'obiettivo di questo lavoro è proporre un indice per valutare la robustezza dei modelli di simulazione e testarne la stabilità

## Materiali e metodi

L'indicatore di robustezza ( $R$ ) è calcolato in base alla seguente equazione  $R = \sigma_{EF} / \sigma_{SAM}$  dove  $\sigma_{EF}$  è la deviazione standard dell'efficienza di modellizzazione ( $EF$ ; Greenwood et al., 1985;  $-\infty \div +1$ ; valore ottimale = 1; se positivo indica che il modello è un miglior stimatore dei dati misurati rispetto alla media dei dati misurati stessi):  $EF = 1 - \left[ \frac{\sum_{i=1}^n (D_i)^2}{\sum_{i=1}^n (M_i - \bar{M})^2} \right]$  dove  $D_i$  è

la differenza tra  $S_i$  e  $M_i$ , con  $S_i$  e  $M_i$  corrispondenti, rispettivamente, all' $i$ -esimo dato simulato e all' $i$ -esimo dato misurato;  $n$  è il numero di coppie  $S_i$ - $M_i$ ;  $\bar{S}$  e  $\bar{M}$  sono le medie dei valori simulati e misurati.

$\sigma_{SAM}$  è la deviazione standard di un indice agrometeorologico sintetico (Synthetic AgroMeteorological indicator,  $SAM$ ,  $-1 \div +1$ ), calcolato con la formula  $SAM = (ET0 - Rain) / (ET0 + Rain)$ , dove  $ET0$  (mm) è l'evapotraspirazione di riferimento

cumula nel periodo 1 marzo – 31 ottobre e  $Rain$  (mm) è la somma delle precipitazioni nello stesso periodo.

$R$  assume valori tra 0 a  $+\infty$ , con valore ottimale = 0.

Al fine di valutare la stabilità dell'indice, abbiamo comparato, entro anno ed entro località, gli ordini assegnati ai valori di  $R$  calcolati per ogni modello utilizzando il top-down concordance coefficient ( $TDCC$ , Iman and Conover, 1987).  $TDCC$  è particolarmente adatto ad enfatizzare concordanze tra gli ordini assegnati a valori importanti (bassi valori di  $R$  in questo caso) e a de-enfatizzare discordanze tra gli ordini assegnati a valori poco importanti (alti valori di  $R$ ).

Indichiamo con  $R_{ij}$  l'indice di robustezza calcolato per il modello  $M_i$  e il dataset  $DS_j$ , e con  $r(R_{ij})$  (con  $i = 1, 2, \dots, k$ ) gli ordini assegnati al valore di  $R$  associato al dataset  $DS_j$ . Un ordine pari a 1 è assegnato al modello  $M_i$  con il valore più basso di  $R_{ij}$ , ordine 2 è assegnato al modello  $M_i$  con il secondo valore più basso di  $R_{ij}$  e così via. Successivamente, vengono calcolati i Savage scores

( $ss(R_{ij})$ ; Savage, 1956) come:  $ss(R_{ij}) = \sum_{i=r(R_{ij})}^k 1/i$ . Per tutti

i modelli  $M_i$  e per tutti i dataset  $DS_j$ .  $TDCC$  è quindi calcolato usando la formula seguente:

$$TDCC = \left\{ \sum_{i=1}^k \left[ \sum_{j=1}^{n_{DS}} ss(R_{ij}) \right]^2 - n_{DS}^2 k \right\} / \left[ n_{DS}^2 \left( k - \sum_{i=1}^k \frac{1}{i} \right) \right]$$

dove  $n_{SD}$  è il numero di dataset. Il  $p$ -value associato ad ogni  $TDCC$  può essere calcolato utilizzando la statistica  $T$  (che approssima una distribuzione  $\chi^2$  con  $k-1$  gradi di libertà), calcolata come:  $T = n_{DS} \cdot (k-1) \cdot TDCC$ .

La concordanza tra due ordini è considerata non statisticamente significativa per  $p$ -value maggiori di 0.05.  $R$  è stato testato (i) su 16 modelli di simulazione raggruppati, sulla base dei processi simulati, in tre diverse categorie, (ii) utilizzando dati raccolti in 13 località dal 1990 al 2007 (Tab. 1).

## Risultati

Tabella 1 mostra i valori di  $R$  calcolati per le tre categorie di modelli esaminate. Tra i modelli per la stima dell'umidità relativa oraria ( $RHh$ ), è possibile notare come RH7 risulti il più robusto, avendo fornito i valori più omogenei di  $EF$  in rapporto alla variabilità dei dataset. Tra i modelli di crescita, WARM ha dimostrato la maggior robustezza, grazie all'elevata stabilità di accuratezza con il quale ha riprodotto curve di crescita misurate in diverse condizioni agrometeorologiche. È interessante notare come WOFOST, il modello più complesso in base al criterio di complessità di Akaike (Confalonieri et al., 2009) risulti il meno robusto. Questo può essere spiegato dal fatto che un modello più complesso del necessario e over-parametrizzato presenti un rischio maggiore di inclusione degli effetti dovuti ad un'annata o ad una località nel valore di alcuni parametri che invece dovrebbero rappresentare solo caratteristiche morfo-fisiologiche della coltura. In questi casi, tipicamente, il modello risulterà inadeguato in condizioni agrometeorologiche diverse da quelle per il quale è stato parametrizzato. Le maggiori difficoltà di CropSyst nel simulare il contenuto di azoto ammoniacale nel suolo dimostrate da Confalonieri et al. (2006) si rispecchiano in una minore robustezza del modello nel simulare la forma ridotta di azoto. Tabella 2 mostra la stabilità di  $R$  – nei confronti dei fattori anno e località – nel momento in cui lo si usa per classificare modelli per la simulazione dello stesso processo biofisico. Nel caso dei 13 modelli per la stima di  $RHh$ , le differenze tra gli ordini assegnati ai modelli non risultano mai statisticamente significative:  $R$  valuta in modo coerente i modelli sia entro anno che entro località. Nel caso dei modelli di crescita colturale,  $TDCC$  dimostra come  $R$  ordini in modo meno stabile i modelli se applicato per valutare la robustezza entro anno o entro località. Questo è in buona parte dovuto al basso numero di modelli testati (3 contro 13 modelli di  $RHh$ ) e al fatto che, su tre modelli, due presentano un valore molto simile dell' $R$  calcolato su tutti i dataset (Tab. 1).

## Conclusioni

Valutare la robustezza dei modelli di simulazione al variare delle condizioni agrometeorologiche nelle quali i modelli vengono applicati è fondamentale per avere garanzie sulle loro performance al di fuori delle condizioni per le quali sono stati parametrizzati, soprattutto per applicazioni a scala territoriale. L'indice proposto si è rivelato utile per fornire una valutazione quantitativa della robustezza, permettendo di classificare diversi approcci di modellazione in base a questo importante criterio e aprendo nuovi scenari per la valutazione integrata di modelli di simulazione.

## Bibliografia

Allen, R.G. et al., 1998. *Irr. & Drain. Paper 56. UN-FAO, Rome, Italy.*  
 ASAE Standards. 1998. EP406.2: heating, cooling, and ventilating greenhouses. St. Joseph, MI, USA.  
 Bekele, D. et al. 2007. Accuracy evaluation of weather data generation and disaggr. methods at finer timescales *Adv W Res* 30, 1286-1300.  
 Confalonieri, R. et al. 2006. The CropSyst model to simulate the N balance of rice for alternative management. *Agr Sust Dev* 26, 241-249.

Tab.1 - Modelli e dataset utilizzati per testare l'indice di robustezza ( $R$ )

Categoria di modello	Variabile simulata	Modello		Dataset		$R$
		Nome	Riferimento	Località	Anni	
Meteo- rologico	Umidità relativa oraria	RH1	a, b, c			0.98
		RH2	a, b, c			1.00
		RH3	a, b			1.67
		RH4	a, b, d			1.23
		RH5	a, b, e		2003,	0.80
		RH6	a, b, f	1, 2, 3,	2004,	1.64
		RH7	g	4, 5, 6,	2005,	0.40
		RH8-0		7, 8, 9	2006,	1.68
		RH8-1			2007	1.36
		RH8-2				1.20
		RH8-3	g, h			1.46
		RH8-4				2.22
		RH8-5				3.45
Crescita colturale	Biomassa aerea (riso)	WARM	i	10, 11,	1990,	0.15
		CropSyst	l	12	1996,	0.30
		WOFOST	m		2002	0.33
Bilancio dell'azoto	Soil N-NO <sub>3</sub> <sup>-</sup> Soil N-NH <sub>4</sub> <sup>+</sup>	CropSyst	l	10, 11	2002,	8.31
					2004	12.30

a: Allen et al., 1998  
 b: ASAE, 1998  
 c: Bekele et al., 2007  
 d: Hubbard et al., 2003  
 e: Linacre, 1992  
 f: Ephrat et al., 1996  
 g: Waichler et al., 2003  
 h: Kimberley et al., 2008  
 i: Confalonieri et al., 2009  
 l: Stöckle et al., 2003  
 m: van Keulen e Wolf, 1986

Tab.2 – Top-down concordance coefficient (TDCC) e relativi  $p$ -value ottenuti dal confronto tra ordini dei modelli sulla base del valore di  $R$  per fattori di variazione località e anno

Categoria di modello	Variabile simulata	TDCC	$p$ -value
<i>Concordanza tra gli ordini ottenuti sui valori di R calcolati per diverse località entro anno</i>			
Meteorologico	Umidità relativa oraria	51.85	6.59E-07
Crescita colturale	Biomassa aerea (riso)	1.14	0.56 *
<i>Concordanza tra gli ordini ottenuti sui valori di R calcolati per diversi anni entro località</i>			
Meteorologico	Umidità relativa oraria	41.29	4.38E-05
Crescita colturale	Biomassa aerea (riso)	0.86	0.65 *

\*: la concordanza tra gli ordini non è statisticamente significativa

Confalonieri, R. et al., 2009. Multi-metric evaluation of the models WARM, CropSyst, and WOFOST for rice. *Ecol. Mod.* 220, 1395-1410.  
 Ephrat, J.E. et al., 1996. Modelling diurnal patterns of air temperatures, radiation, wind speed and relative humidity by equations for daily characteristics. *Agricultural Systems*, 51:377-393.  
 Fila, G. et al. 2003. IRENE: a software to evaluate model performance. *Eur. J. Agron.*, 18, 369-372.  
 Greenwood, D.J. et al. 1985. Response of potatoes to N fertilizer: dynamic model. *Plant Soil*, 85, 185-203.  
 Hubbard, K.G. et al. 2003. Estimating daily dew point temp. for the North. *Great Plains using max. and min. temp. A. Jour.* 95 323-328.  
 Iman, R.L., Conover, W.J., 1987. A measure of top-down correlation. *Technometrics* 29, 351-357.  
 Kimberley, R., Center, E., 2008. ASCE-EWRI Standardization of reference evapotranspiration: draft report appendices <http://www.kimberly.uidaho.edu/water/asceewri/index>.  
 Linacre, E. 1992. *Climate data and resources: a reference and guide.* Routledge, London, United Kingdom.  
 Savage, I. R., 1956. Contributions to the Theory of Rank Order Statistics – the Two-sample Case. *Ann. Math. Statist.* 27, 590-615.  
 Stöckle, C.O. et al. 2003. CropSyst, a cropping systems simulation model. *Eur. J. Agron.* 18, 289–307.  
 Van Keulen, H., Wolf, J., 1986. Modelling of agricultural production: weather soils and crops. *Simul. Monographs*, Pudoc, Wageningen.  
 Waichler, S. R., Wigmosta, M. S., 2003. Development of hourly meteor. values from daily data and significance to hydrological modeling at H. J. Andrews Experimental Forest. *J. Hydromet.* 4, 251-263.