Fuzzy-logic based multi-site crop model evaluation in europe

R. Ferrise[1] – M. Bindi[1] – M. Acutis[2] – G. Bellocchi[3]

[1] DISPAA, University of Florence, P.le delle Cascine 18, Florence, Italy, roberto.ferrise@unifi.it
[2] DiSAA, University of Milan, Via G. Celoria 2, Italy, marco.acutis@unimi.it
[3] INRA-UREP, French National Institute for Agricultural Research-Grassland Ecosystem Research Unit, 5 Chemin de Beaulieu, Clermont-Ferrand, France, gianni.bellocchi@clermont.inra.fr
Corresponding author: roberto.ferrise@unifi.it

Introduction

It is agreed that the aggregation of metrics of different nature into integrated indicators offers a valuable way to assess models (Bellocchi et al., 2010). A composite indicator (MQI_m: Model Quality Indicator for multi-site assessment) was elaborated on metrics commonly used to evaluate simulation models, and on recent concepts of model evaluation integrating sensitivity analysis and robustness measures, as well as information criteria for model selection and expert judgments on the importance of different metrics. For wheat modelling in Europe, we document to what extent the MQI_m reflects the main components of model quality and supports inferences about model performances.

Materials and Methods

Five crop simulation models (CropSyst, DSSAT, HERMES, SIMPLACE, STICS), differing in processes and approaches used to represent the dynamics of crop phenology and growth, were applied to reproduce winter wheat development, aboveground biomass and yield at three experimental sites in Europe (Tab. 1).

Table 1. Sites and experimental setup (Kollas et al., 2015).

Site characteristics	Foulum	Müncheberg	Thibie
Country	Denmark	Germany	France
Latitude (decimal degrees North)	56.49	52.52	48.93
Longitude (decimal degrees East)	9.57	14.12	4.23
Climate type*	Atlantic North	Continental	Atlantic Central
Years of available data	2003-2012	1993-1998	1992-2003

* Metzger et al., (2005).

The performance of models was assessed using the MQI_m (http://ojs.macsur.eu/index.php/Reports/article/view/D-L2.2/59), which aggregates three components (modules) of model quality: agreement with actual data, complexity of the model, and stability of performance over a range of conditions (robustness). Using fuzzy logic-based weighting, a number of basic performance metrics were converted and aggregated into modules, which are dimensionless values between 0 (best model response) and 1 (worst model response). Then, the modules were aggregated in the final indicator (as well, dimensionless and in the range 0 to 1).

Results and Discussion

Limited to simulated grain yield, results show that the ranking of models may change depending on the metric considered (Tab. 2), thus confirming the need to aggregate several metrics to have a comprehensive view of model performance. MQI_m values >0.7 with all the models indicate that wheat yield is difficult to simulate, with high variability depending on the site (Robustness=1). Under the conditions evaluated, model 3 resulted the best-performing. Its performance was mainly due to a better agreement to data (d=0.80), achieved thanks to its high complexity (Complexity=1) owing to a relative high proportion of relevant parameters (R_p>0.5).

Table 2. Winter wheat yield simulations: performance metrics, modules and MQI_m calculated over three sites with five crop models. I_R: index of robustness (best, 0 ÷ ∞, worst); P(t): probability of paired Student t-test for means being equal (worst, 0 ÷ ∞, best), d: index of agreement (worst, 0 ÷ 1, best), R: Pearson's correlation coefficient of the estimates versus measurements (worst, -1 ÷ +1, best), R_p: relevant parameter ratio (best, 0 ÷ 1, worst), w_k: ratio of Akaike's Information Criterion (worst, 0 ÷ 1, best). For each basic metric, the upper line indicates the average value across sites. Greyed areas show the best value per metric.

Model	Performance metrics, modules and indicator					
	$\overline{P(t)}$	$\bar{r}$	$\bar{d}$	$\overline{R_p}$	$\overline{w_k}$	I_R
1	0.22	0.30	0.50	0.32	2.10E-13	31.7
2	0.22	0.30	0.57	0.28	3.26E-10	66.7
3	0.42	-0.04	0.80	0.53	0.24	47.2
4	0.26	-0.05	0.27	0.50	0.76	176.7
5	0.29	0.20	0.43	0.37	1.07E-08	211.9
	Agreement			Complexity		Robustness
1	0.8182			0.7975		1.000
2	0.8182			0.7975		1.000
3	0.5000			1.0000		1.000
4	0.8000			0.5029		1.000
5	0.8000			0.8944		1.000
	MQI_m					
1	0.9128					
2	0.9128					
3	0.7500					
4	0.8060					
5	0.9504					

Conclusions

The aggregation of different aspects of model quality into a single indicator allowed ranking models from the best to the worst-performing. However, this is not a conclusive judging about the quality of models evaluated (kept anonymous). Rather, it is meant to help modellers identifying areas of their model requiring improvement.

Acknowledgements

Work supported by the FACCE JPI projects MACSUR and CN-MIP, and the EU-FP7 project MODEXTREME.

References

Bellocchi G, Rivington M, Donatelli M, Matthews K (2010) Agron Sustain Dev 30:109-130
Kollas C, Kersebaum KC, Nendel C, et al., (2015) Eur J Agronomy 70:98-111
Metzger MJ, Bunce RGH, Jongman RHG, Mucher CA, Watkins JW (2005) Global Ecol Biogeogr 14:549-563