

Crop simulation models evaluation

Gianni Bellocchi

▶ **To cite this version:**

Gianni Bellocchi. Crop simulation models evaluation. Master. Master science (Evaluation de modèles), 2015, 23 p. hal-02792954

HAL Id: hal-02792954
https://hal.inrae.fr/hal-02792954

Submitted on 5 Jun 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Crop simulation models evaluation

Gianni BELLOCCHI

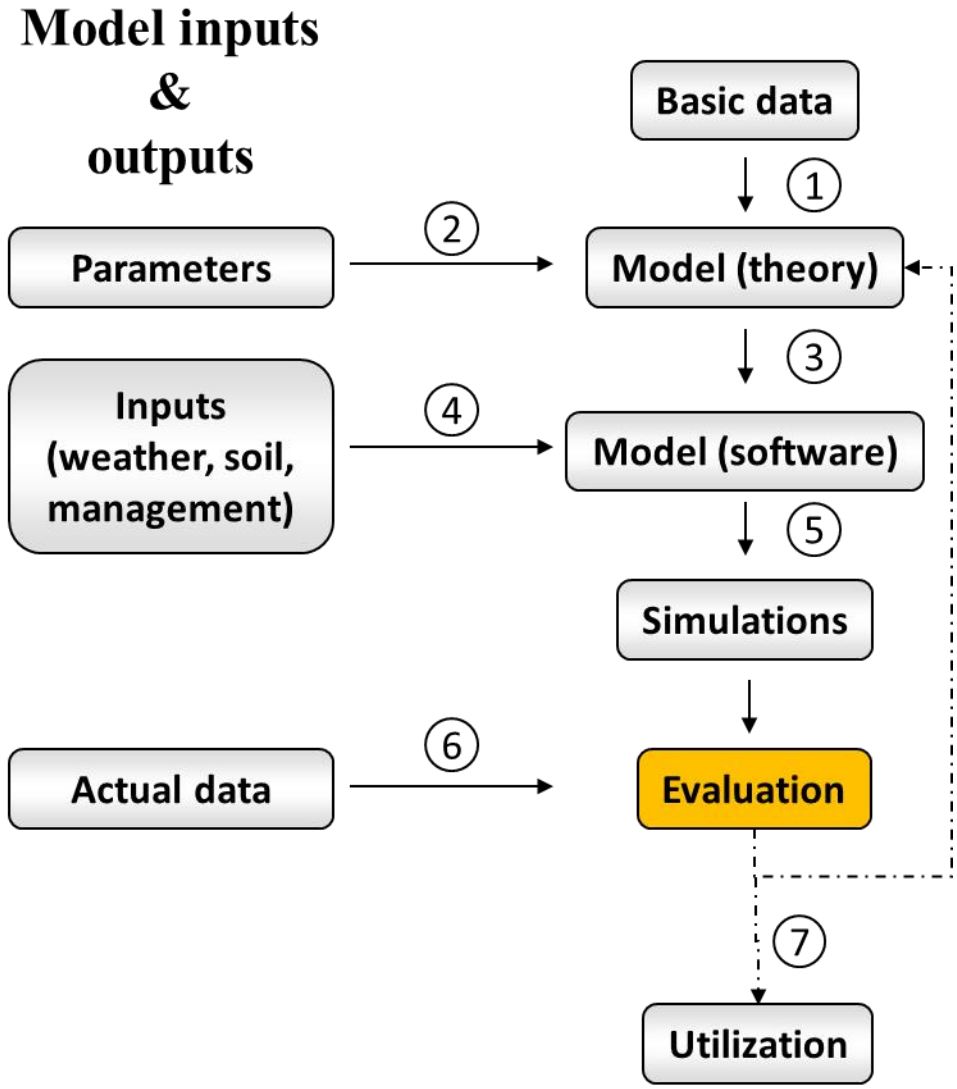
Grassland Research Ecosystem Unit - French National Institute for Agricultural Research
Clermont-Ferrand, France

28 April 2015
Sassari (Italy)

Agroecological system and modelling process



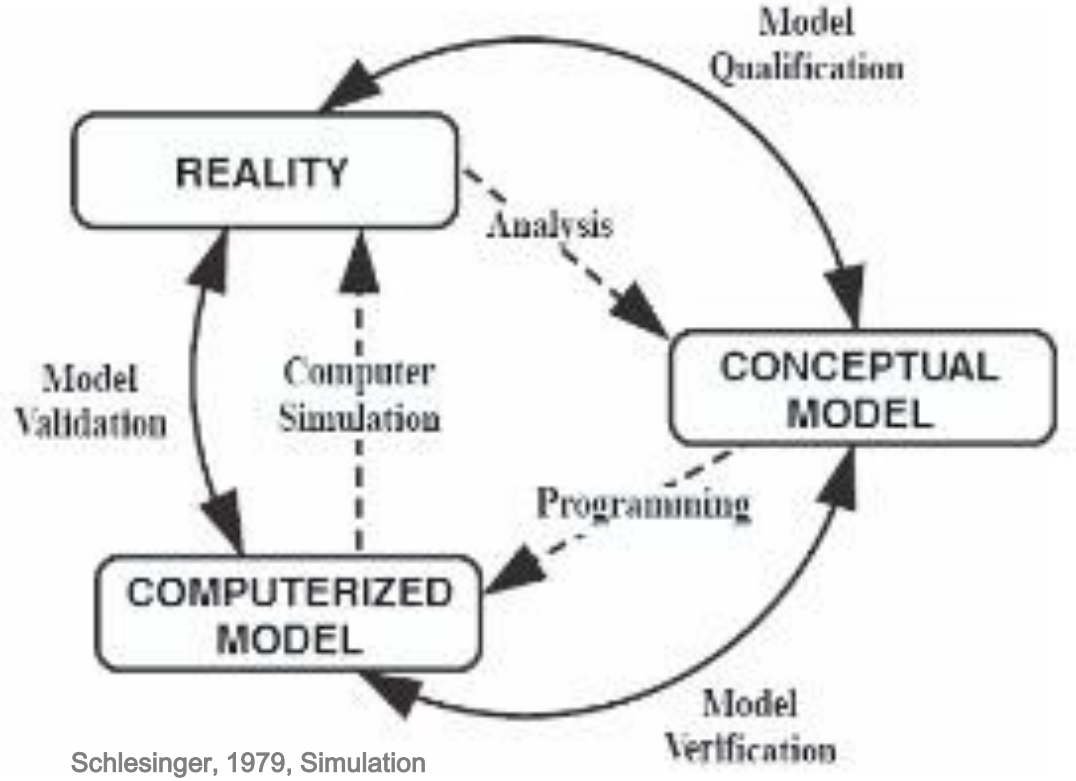
Actual system



Modelling flow

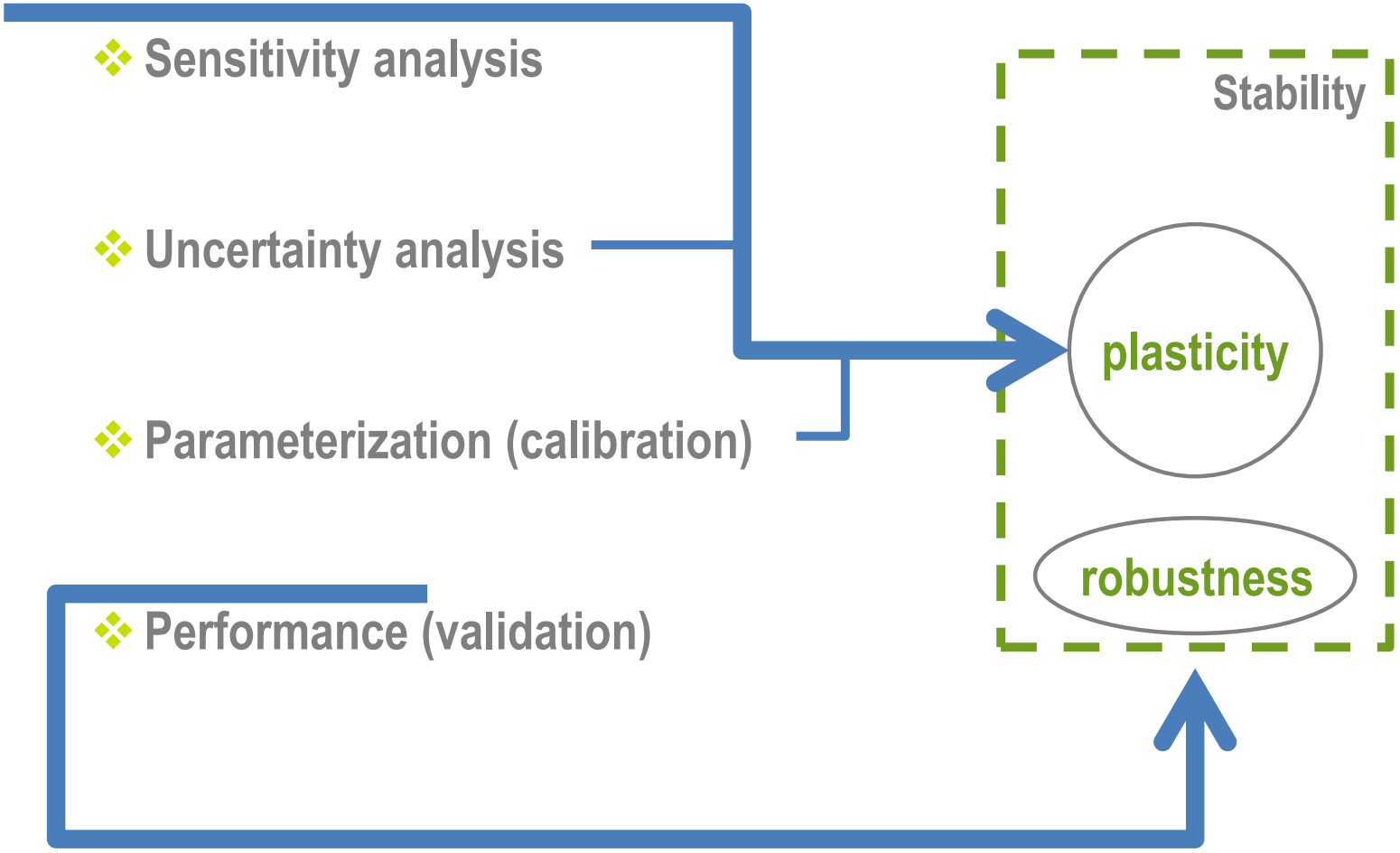
Model evaluation

Model evaluation **(validation + verification)**:
action in which the quality of a mathematical model for specific objectives is established

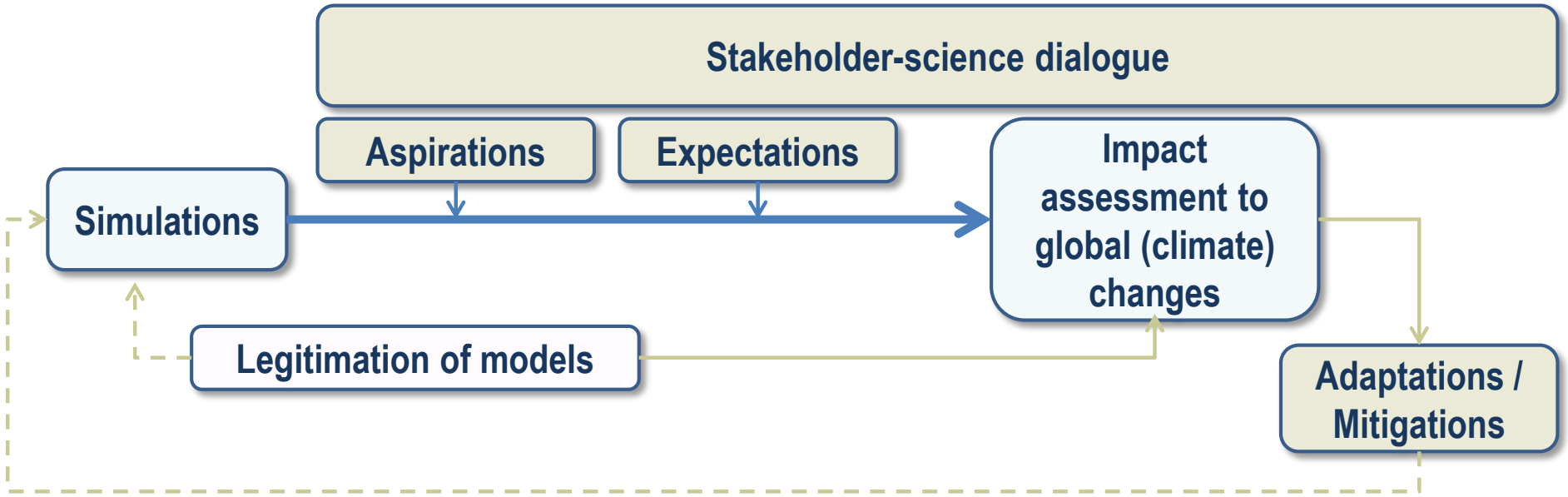


Schlesinger, 1979, Simulation

Model evaluation components



Deliberative process in model-based climate change studies



Bellocchi et al. (2006)

Rivington et al. (2007)

Bellocchi et al. (2015)

Ital. J. Agron., 3, 1, 19-39

A Review of Methodologies to Evaluate Agroecosystem Simulation Models

F. MARTORANA and G. BELLOCCHI

Agron. Sustain. Dev. (2009)
© INRA, EDP Sciences, 2009
DOI: 10.1051/agro/2009001

Available online at:
www.agronomy-journal.org

Review article

Validation of biophysical models: issues and methodologies. A review

Gianni BELLOCCHI[1*,**], Mike RIVINGTON[2], Marcello DONATELLI[1***], Kees

Environmental Modelling & Software 26 (2011) 328–336

Contents lists available at ScienceDirect

Environmental Modelling & Software

journal homepage: www.elsevier.com/locate/envsoft

Technical assessment and evaluation of environmental models and software: Letter to the Editor

Alexandrov[a,*], D. Ames[b], G. Bellocchi[c], M. Bruen[d], N. Crout[e], M. Erechtchoukova[f], A. Hildebrandt[g], Matsunaga[a], S.T. Purucker[k], M. Rivington[l],

Contents lists available at Science

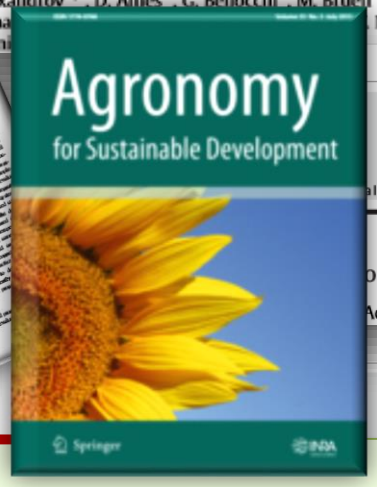
Agricultural Systems

journal homepage: www.elsevier.com/loc

An evaluation of the statistical methods for testing the pe of crop models with observed data

J.M. Yang[a], J.Y. Yang[b,*], S. Liu[b,c], G. Hoogenboom[d]





Ecological Modelling 220 (2009) 1395–1410

Contents lists available at ScienceDirect

Ecological Modelling

journal homepage: www.elsevier.com/locate/ecolmodel

of the models WARM, CropSyst, and WOFOST for rice

Acutis[b], Gianni Bellocchi[c], Marcello Donatelli[d,1]

Ecological Modelling 221 (2010) 960–964

Contents lists available at ScienceDirect

Ecological Modelling

/locate/ecolmodel

Assessment dequacy of

Journal of
Applied Remote Sens

obustness based on the
xplored conditions

Spanish Journal of Agricultural Research 2009 7(3), 680-686
ISSN: 1695-971-X

strategies for rice modelling

M. Boschetti[3] and M. Acutis[4]

Derivation of biophysical variables Earth observation data: validation statistical measures

Katja Richter
Clement Atzberger
Tobias B. Hank
Wolfram Mauser

An Indi

y Expert System

li

Elaboration of new metrics

Setting of thresholds

Meaning and limitations

Intercorrelation

Disaggregation

Aggregation

Some metrics

$$CRM = \frac{\sum_{i=1}^{n} O_i - \sum_{i=1}^{n} P_i}{\sum_{i=1}^{n} O_i}$$

$$MAE = \frac{\sum_{i=1}^{n} |P_i - O_i|}{n}$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n} (P_i - O_i)^2}{n}}$$

simple

absolute

squared

difference-based metrics

non-parametric

$$MdAE = \text{median}_{i=1,\ldots,n} |P_i - O_i|$$

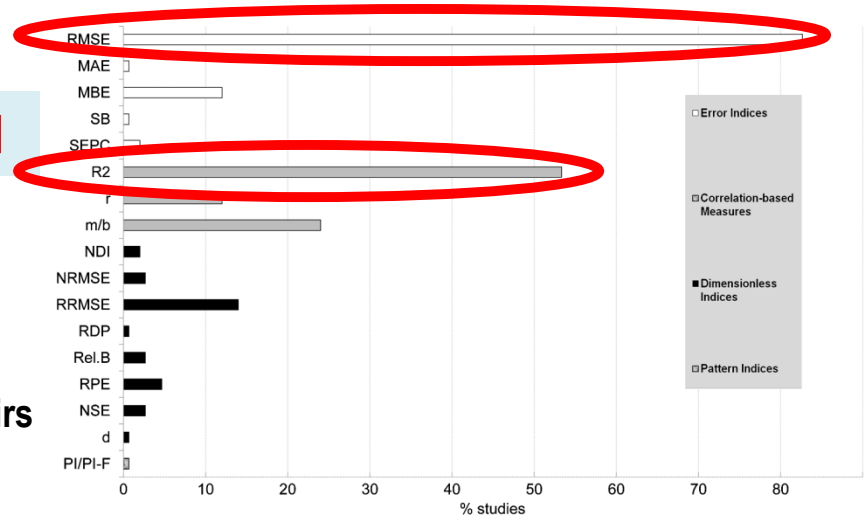
association-based metrics

$$r = \frac{\sum_{i=1}^{n} (P_i - \overline{P}) \cdot (O - \overline{O})}{\sqrt{\sum_{i=1}^{n} (P_i - \overline{P})^2 \cdot \sum_{i=1}^{n} (O_i - \overline{O})^2}} \qquad r^2, slope, intercept$$

combined

$$EF = 1 - \frac{\sum_{i=1}^{n} (P_i - O_i)^2}{\sum_{i=1}^{n} (O_i - \overline{O})^2}$$

$$d = 1 - \frac{\sum_{i=1}^{n} (P_i - O_i)^2}{\sum_{i=1}^{n} (|P_i - \overline{O}| + |O_i - \overline{O}|)^2}$$



P = predicted; O = observed; i = i^{th} O/P pair; n = number of O/P pairs

Richter et al. (2012)

Setting of thresholds

Performance measure	Unit	Value range and purpose	Reliability criteria
Coefficient of determination ($\mathbf{R^2}$) of the linear regression estimates versus measurements	dimensionless	0 (absence of fit) to 1 (perfect fit): the closer values are to 1, the better the model	> 0.8
Willmott (1982) index of agreement (**d**)	dimensionless	0 (absence of agreement) to 1 (perfect agreement): the closer values are to 1, the better the model	> 0.8
Mean absolute error over the mean of the measured values (**MAE(%)**)	%	0 (optimum) to positive infinity: the smaller MAE(%), the better the model performance	< 20

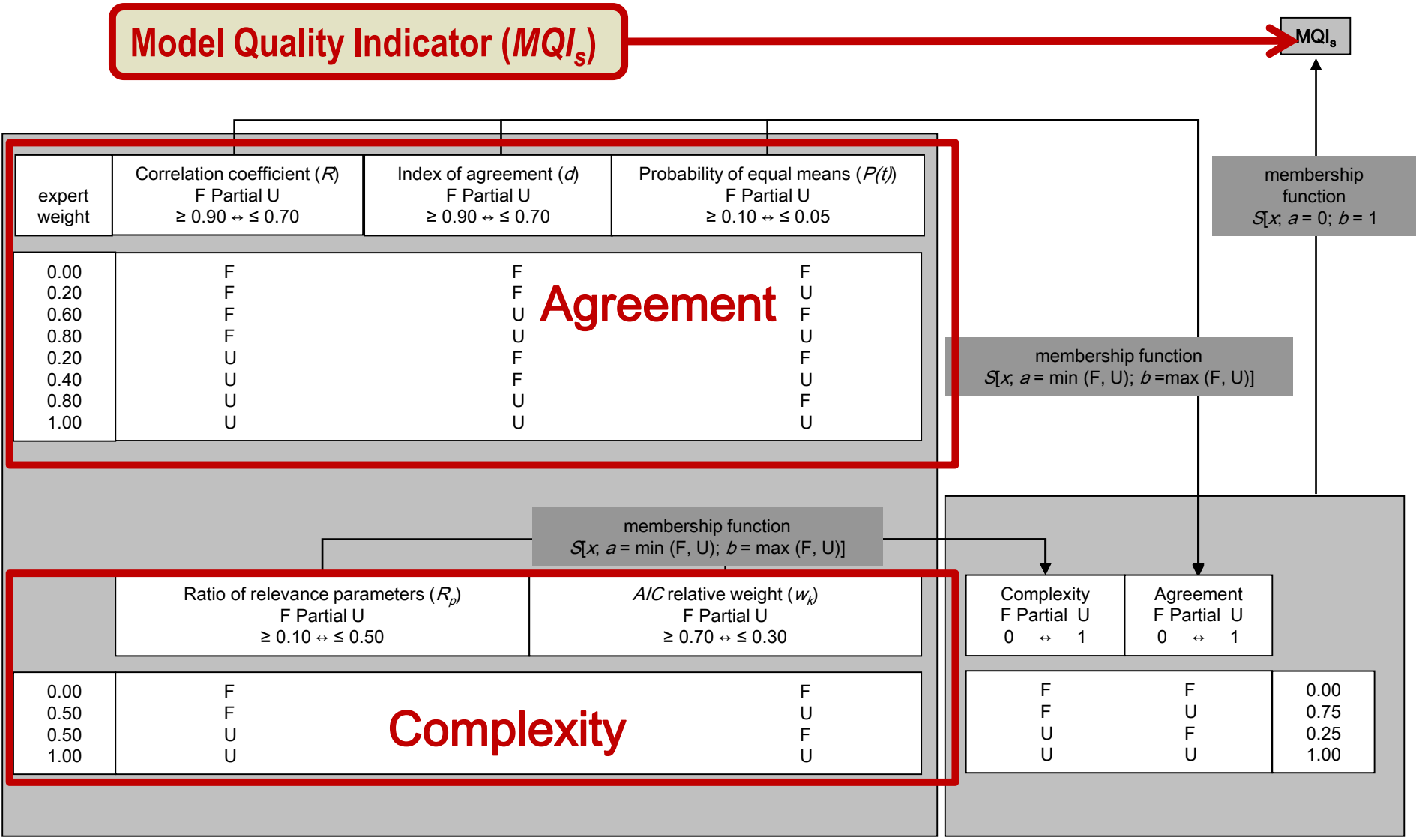
De Jager (1994)

Key issues and factors

Key validation issues	Major factors to investigate				
	Modelling objective	Model inputs	Model outputs	Model structure	Modelling conditions
Validation purpose	X		X		X
Robustness of results			X		X
Interpretation of phenomena		X	X	X	
Model comparison				X	
Model predictions	X		X		X
Model complexity		X	X	X	
Data accuracy		X	X		
Time histories			X		

Bellocchi et al. (2010)

Fuzzy-logic based indicators



Multi-site, Model Quality Indicator (MQI_m)

MQI_m

membership function
$S[x; a = \min(F, U); b = \max(F, U)]$

membership function
$S[x; a = 0; b = 1$

Agreement

expert weight	Correlation coefficient (R) F Partial U ≥ 0.90 ↔ ≤ 0.70	Index of agreement (d) F Partial U ≥ 0.90 ↔ ≤ 0.70	Probability of equal means ($P(t)$) F Partial U ≥ 0.10 ↔ ≤ 0.05
0.00	F	F	F
0.20	F	F	U
0.60	F	U	F
0.80	F	U	U
0.20	U	F	F
0.40	U	F	U
0.80	U	U	F
1.00	U	U	U

membership function
$S[x; a = \min(F, U); b = \max(F, U)]$

Complexity

	Ratio of relevance parameters (R_p) F Partial U ≥ 0.10 ↔ ≤ 0.50	AIC relative weight (w_k) F Partial U ≥ 0.70 ↔ ≤ 0.30
0.00	F	F
0.50	F	U
0.50	U	F
1.00	U	U

Complexity F Partial U 0 ↔ 1	Agreement F Partial U 0 ↔ 1	Robustness F Partial U 0 ↔ 1
0.00	F	F	F
0.25	F	F	U
0.50	F	U	F
0.75	F	U	U
0.25	U	F	F
0.50	U	F	U
0.75	U	U	F
1.00	U	U	U

Robustness

	Index of robustness (I_R) F Partial U 1 ↔ 10
0.00	F
1.00	U

membership function
$S[x; a = \min(F, U); b = \max(F, U)]$

Robustness of a model

A **robustness measure** would account for model performance stability over a wide range of conditions (single site versus multiple sites)

How the variability of model performance can be quantified with the variability of conditions?

Index of robustness

$$I_R = \frac{\sigma_{EF}}{\sigma_{SAM}}$$ (0, best; +∞, worst)

Modelling efficiency

$$EF = 1 - \frac{\sum_{i=1}^{n}(P_i - O_i)^2}{\sum_{i=1}^{n}(O_i - \bar{O})^2}$$ (-∞, worst; 1, best)

Synthetic Agro-Meteorological Indicator

$$SAM = \frac{Rain - ET_0}{Rain + ET_0}$$ (-1, +1)

Confalonieri et al. (2010)

Synthetic indicators

Aggregation rules:
fuzzy-logic based weighing system

I. Agreement
- Correlation coefficient
- Index of agreement
- Probability of equal means

II. Complexity
- Ratio of relevant parameters
- Parameters-agreement criterion

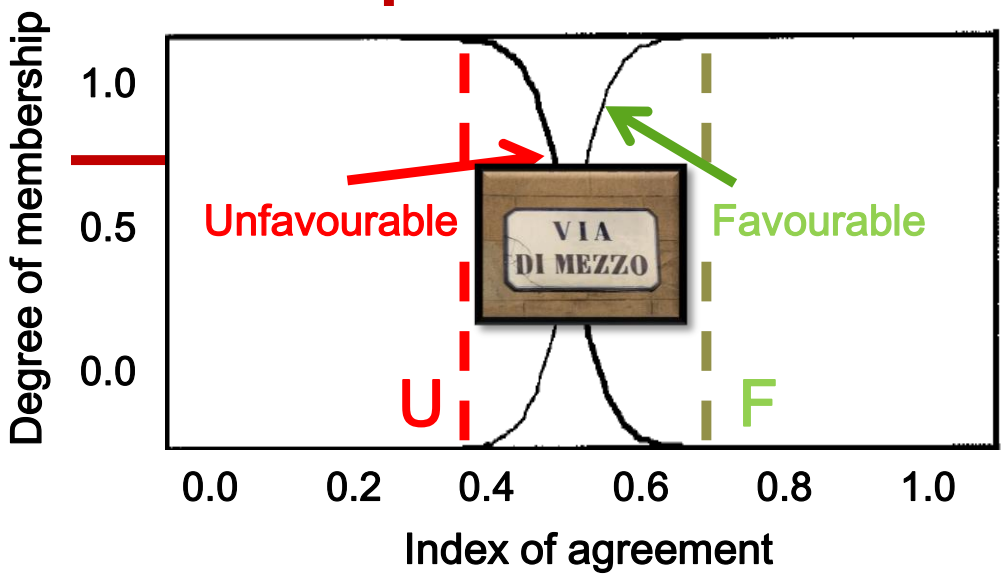
III. Stability (robustness)
- Index of robustness

Non-dimensionality

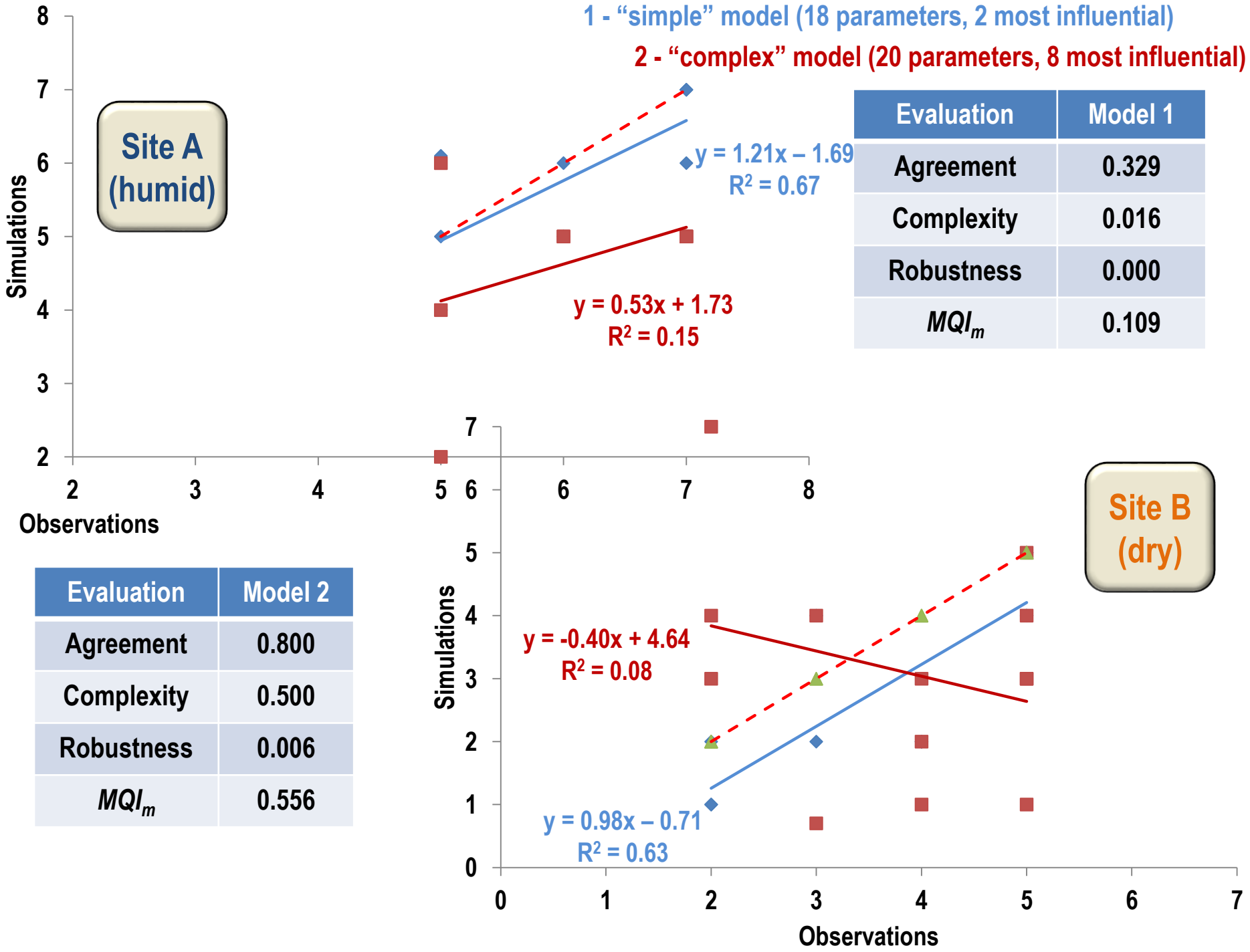
Lower and upper bounding

(best) 0 – 1 (worst)

Model Quality Indicator



Hindrances to overcome:
thresholds and weights



Evaluation	Model 1
Agreement	0.329
Complexity	0.016
Robustness	0.000
MQI_m	0.109

Evaluation	Model 2
Agreement	0.800
Complexity	0.500
Robustness	0.006
MQI_m	0.556



Rice simulations:

above-ground biomass at maturity

Three models: CropSyst (simple), WARM (intermediate), WOFOST (complex)

MQI_s	M1	M2	M3
C. d'Agogna	**0.0313**	0.1250	0.2174
Vercelli	0.1070	**0.0853**	0.1372
Mortara	0.2188	**0.0000**	0.2174
Rosate	**0.0313**	0.2284	0.2388
MQI_m	**0.0750**	0.1940	0.3356

EF	M1	M2	M3
C. d'Agogna	0.90	0.95	0.93
Vercelli	0.92	0.97	0.96
Mortara	0.96	**0.98**	**0.98**
Rosate	0.92	**0.62**	**0.48**
I_R	**0. 16**	1.24	1.71

Robustness

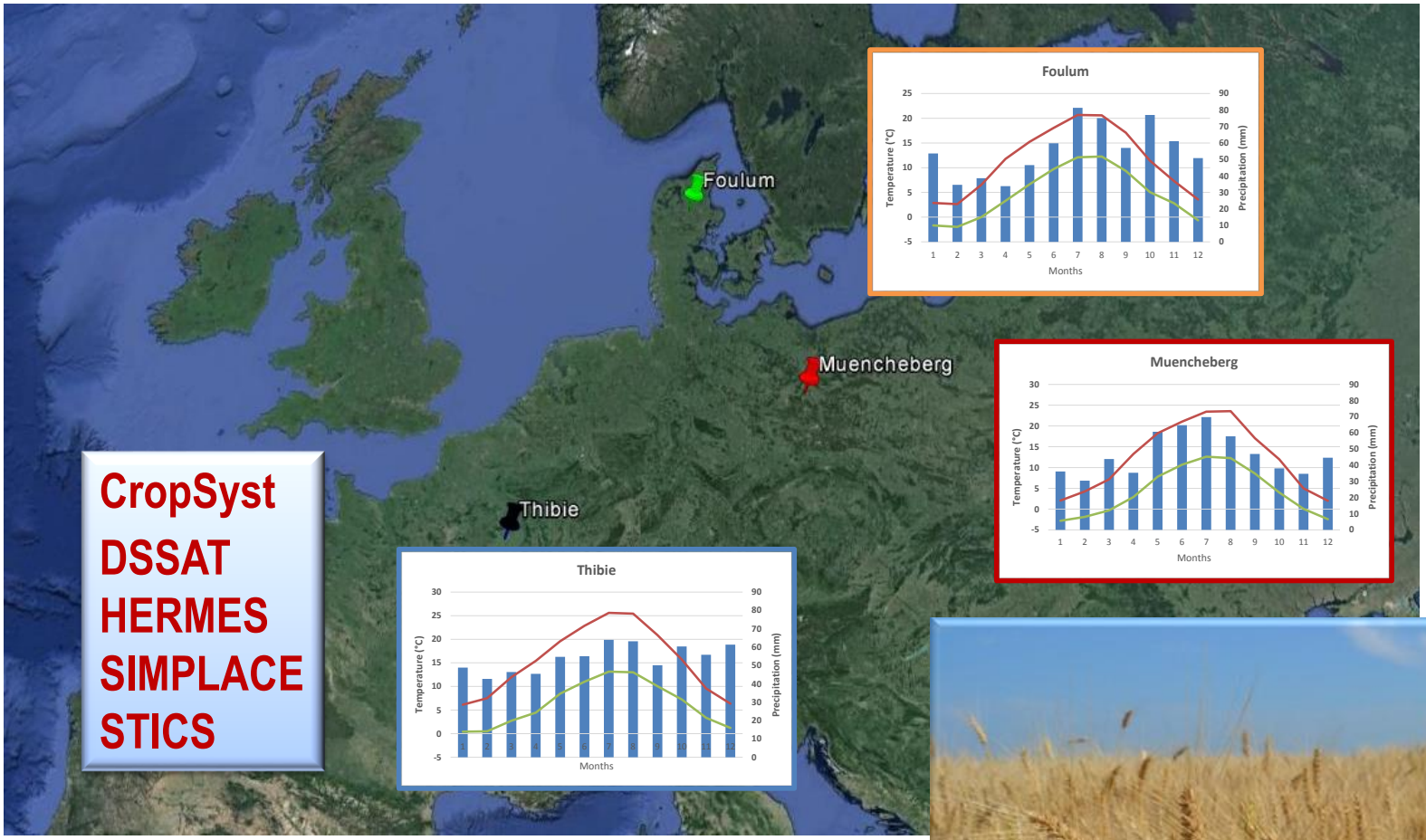
MSE	M1	M2	M3
C. d'Agogna	3.26	**1.86**	2.42
Vercelli	2.93	**1.35**	1.57
Mortara	1.66	**0.84**	0.94
Rosate	**0.97**	**4.96**	**6.75**

AIC	M1	M2	M3
C. d'Agogna	**34**	37	**79**
Vercelli	**33**	34	**73**
Mortara	**26**	28	**67**
Rosate	**20**	49	**91**

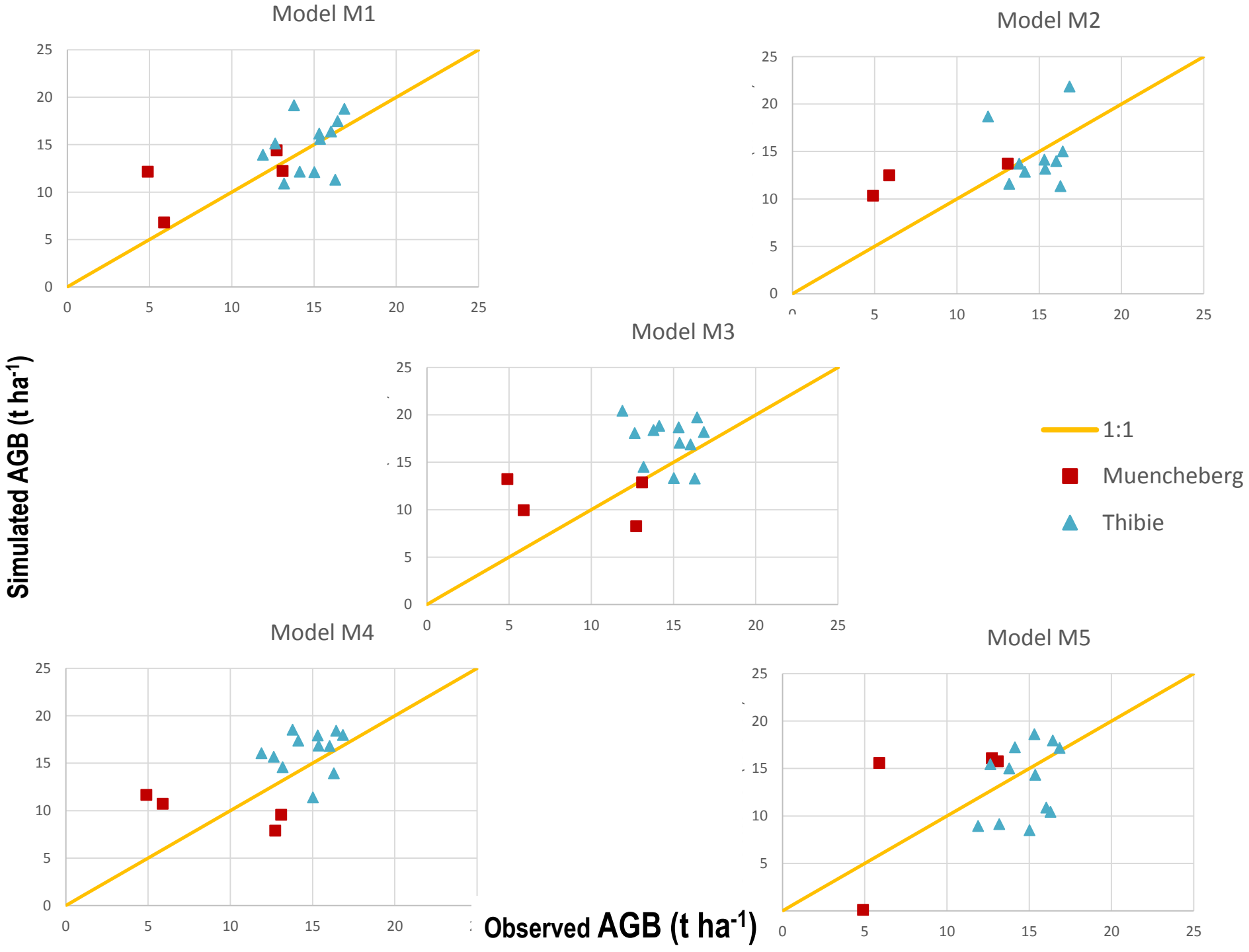
Complexity

Wheat simulations:

above-ground biomass at maturity

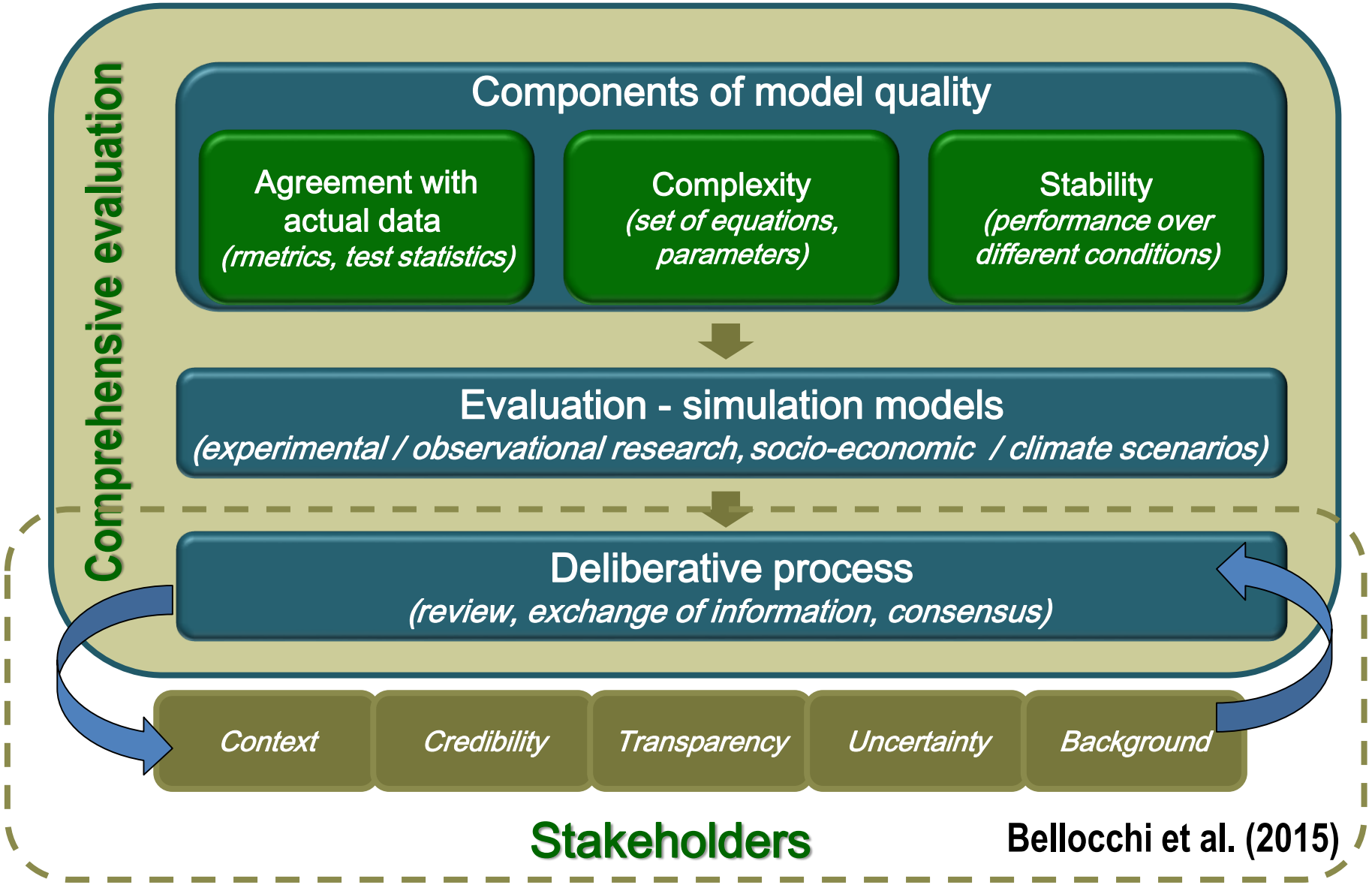





Model M1
Model M2
Model M3
Model M4
Model M5
Simulated AGB (t ha-1)
Observed AGB (t ha-1)
1:1
Muencheberg
Thibie

Model	Aboveground biomass at maturity: performance metrics, modules and indicator					
	$\overline{P(t)}$	$\bar{r}$	$\bar{d}$	$\overline{R_p}$	$\overline{w_k}$	I_R
M1	0.23	0.46	0.64	0.32	1.99E-13	65.4
M2	0.20	0.46	0.60	0.28	2.66E-11	6.0
M3	0.01	-0.25	0.70	0.53	0.12	149.5
M4	0.08	-0.36	0.25	0.50	0.88	344.6
M5	0.08	0.49	0.60	0.37	1.34E-08	377.6
	Agreement			Complexity		Robustness
M1	0.8000			0.7975		1.0000
M2	0.8000			0.7975		0.6049
M3	1.0000			1.0000		1.0000
M4	0.8640			0.5000		1.0000
M5	0.8640			0.8944		1.0000
	MQI_m					
M1	0.8976					
M2	0.7471					
M3	1.0000					
M4	0.8428					
M5	0.9640					

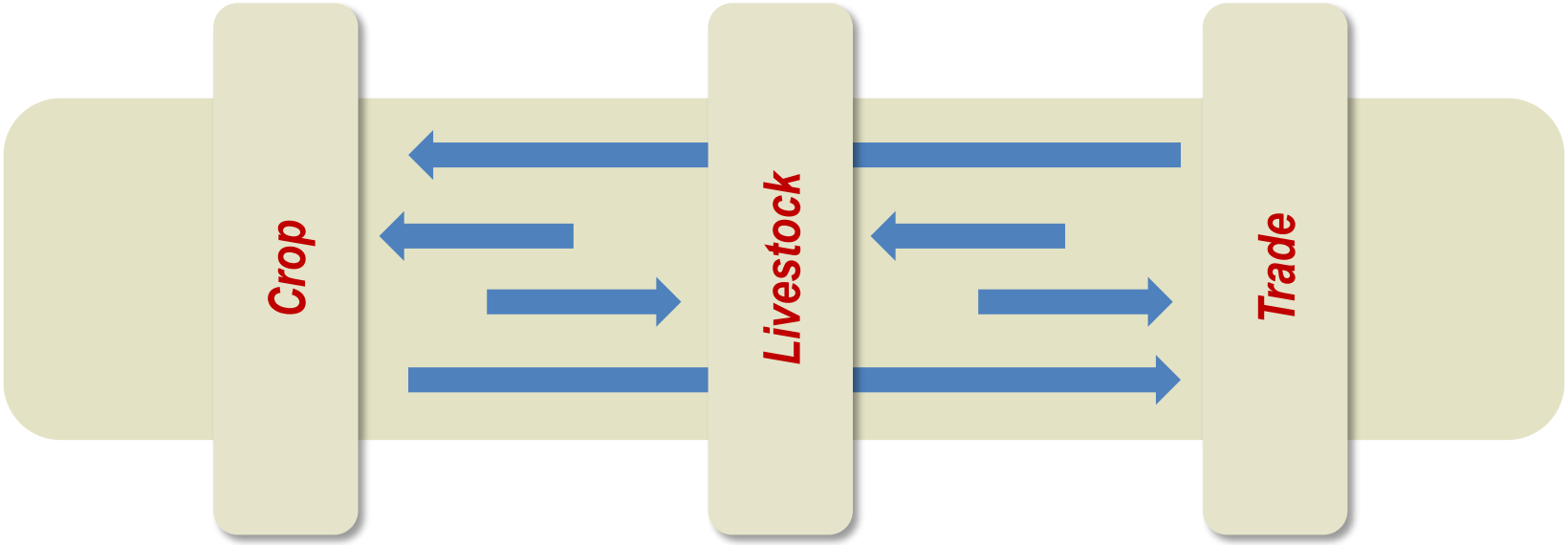
Model evaluation / deliberative process



Bellocchi et al. (2015)

Towards a consolidated, internationally-agreed protocol to evaluate models: what does go forth?

- Review of settings
  - Selection of metrics
  - Attribution of thresholds and weights
- Extension to multiple outputs



Institutionalising deliberative practices for context-specific model evaluations

Model evaluation(s) are (sometimes) an (important) orientating landmark in the skyline of decisions, without replacing them

To evaluate simulation models (in agriculture) is far more urgent as many of the (tactical and strategic) decisions (in agriculture) are based on model outcomes

Dealing with (existing) and designing (new) agricultural systems is a priority that deliberations about model evaluation contribute to accomplish in a more efficient (maybe more appropriate) manner, in any case with more awareness if (genuine) collective deliberations are possible

The central issue is to think and conceive model evaluation in a (clear) decisional perspective about type of model, operability, transparency, etc.

As several models are at hand, "mod-diversity" imposes the analysis of case-by-case issues, while also integrating the specific context in a larger-scale perspective (in space and time)

Literature sources

Bellocchi G., Confalonieri R., Donatelli M., 2006. Crop modelling and validation: integration of IRENE_DLL in the WARM environment. Italian Journal of Agrometeorology 11, 35-39.

Bellocchi G., Rivington M., Matthews K., Acutis M., 2015. Deliberative processes for comprehensive evaluation of agroecological models. A review. Agronomy for Sustainable Development 35, 589-605.

Bellocchi G., Rivington M., Donatelli M., Matthews K.B., 2010. Validation of biophysical models: issues and methodologies. A review. Agronomy for Sustainable Development 30, 109-130.

De Jager J.M., 1994. Accuracy of vegetation evaporation formulae for estimating final wheat yield. Water SA 20, 307-314.

Richter K., Atzberger C., Hank T.B., Mauser W., 2012. Derivation of biophysical variables from Earth observation data: validation and statistical measures. Journal of Applied Remote Sensing 6, 063557.

Rivington M., Matthews K.B., Bellocchi G., Buchan K., Stöckle C.O., Donatelli M., 2007. An integrated assessment approach to conduct analyses of climate change impacts on whole-farm systems. Environmental Modelling & Software 22, 202-210.

Schlesinger S., 1979. Terminology for model credibility. Simulation 32, 103-104


Thank you for your attention.