Deliberative processes for comprehensive evaluation of agro-ecological models

Gianni BELLOCCHI

French National Institute for Agricultural Research, Clermont-Ferrand, France

Mike RIVINGTON

The James Hutton Institute, Aberdeen, United Kingdom

Marco ACUTIS

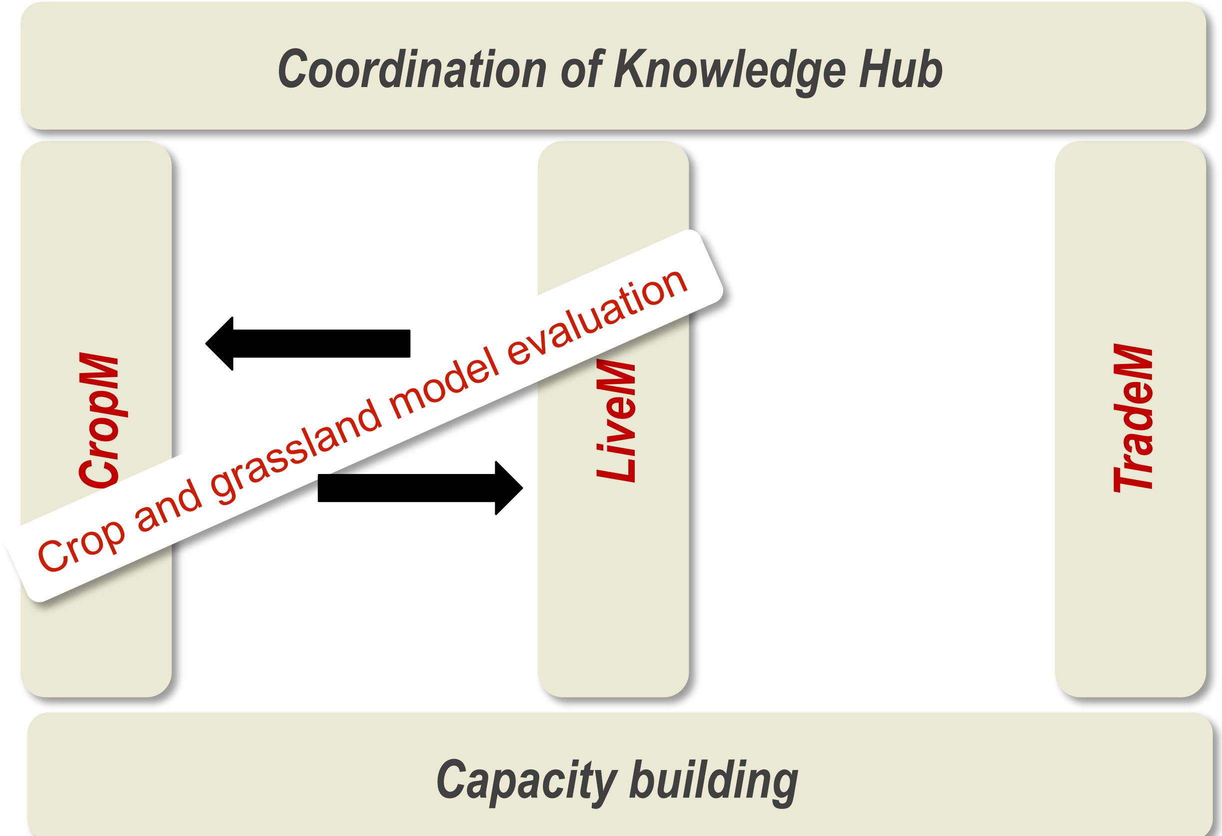
University of Milan, Italy

FACCE MACSUR Mid-Term Scientific Conference

University of Sassari, Italy

01-04 April 2014

MACSUR cross-cutting activities



CropM-LiveM

- *Definition of model performance indicators*
- *Elaboration of model evaluation protocols*

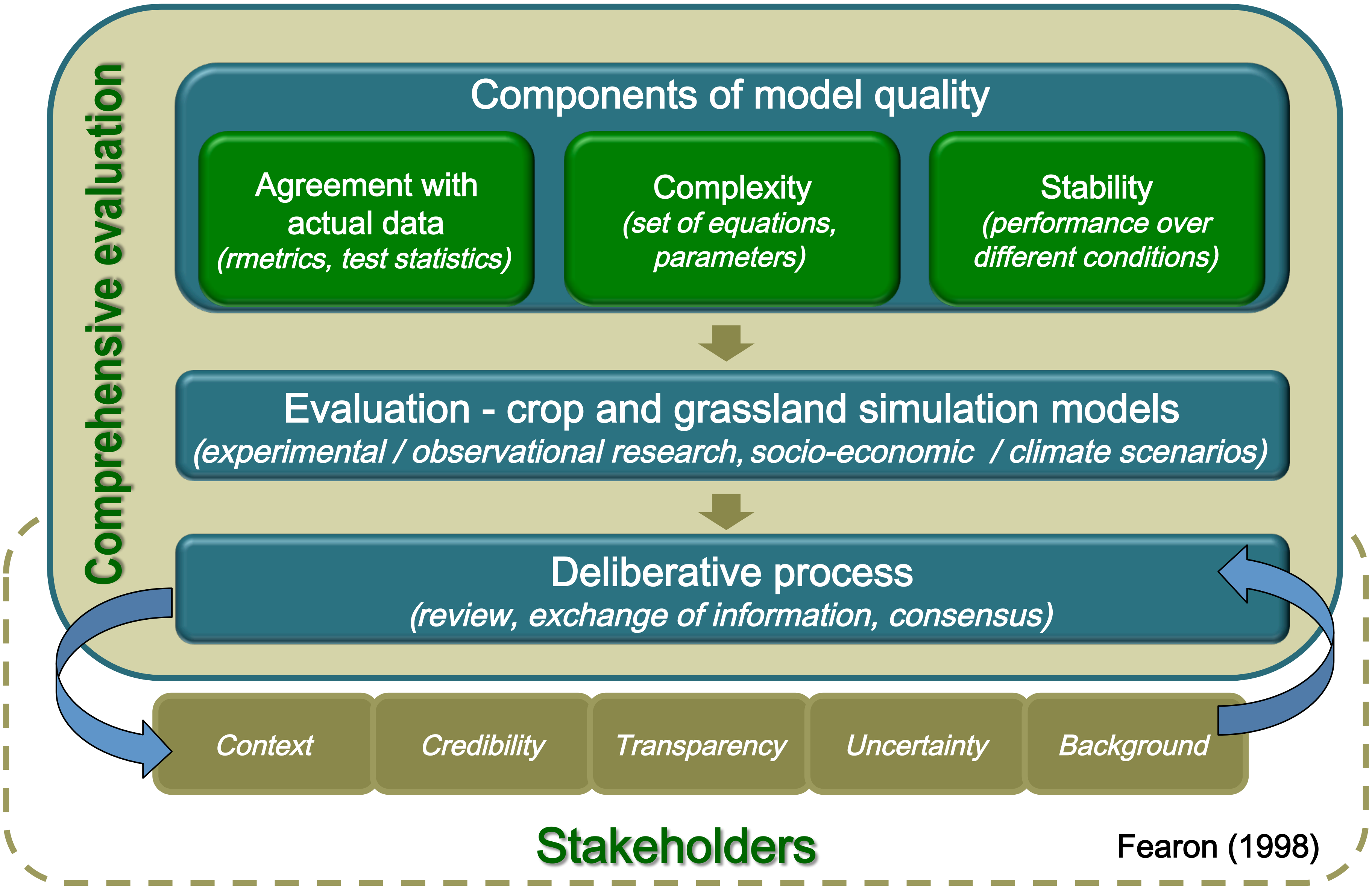
Task C1.4

Develop and apply model evaluation methods

Task L2.2

Development of methods for model evaluation

Model evaluation / deliberative process



Synthetic indicators

Aggregation rules: fuzzy-logic based weighing system

Non-dimensionality

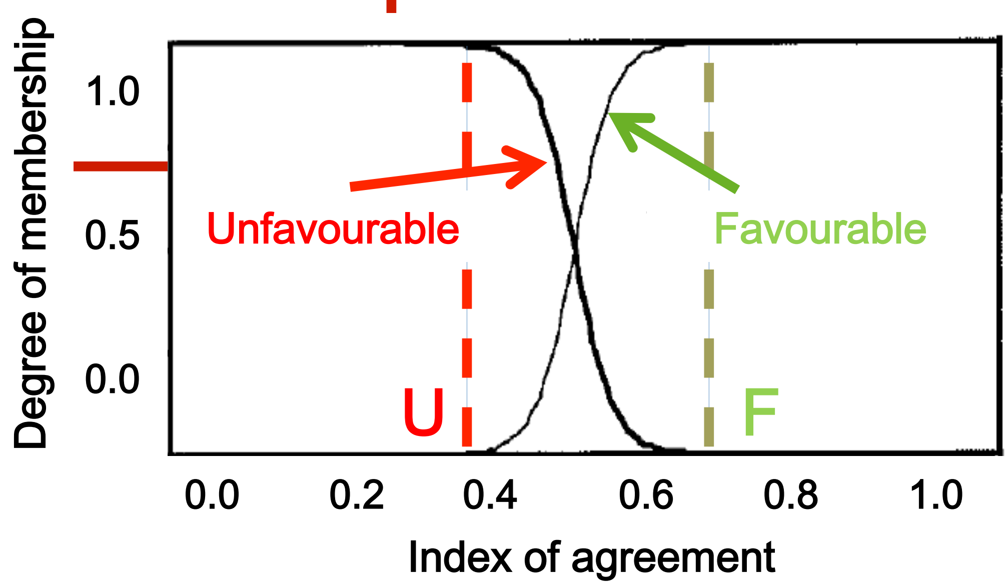
Lower and upper bounding

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

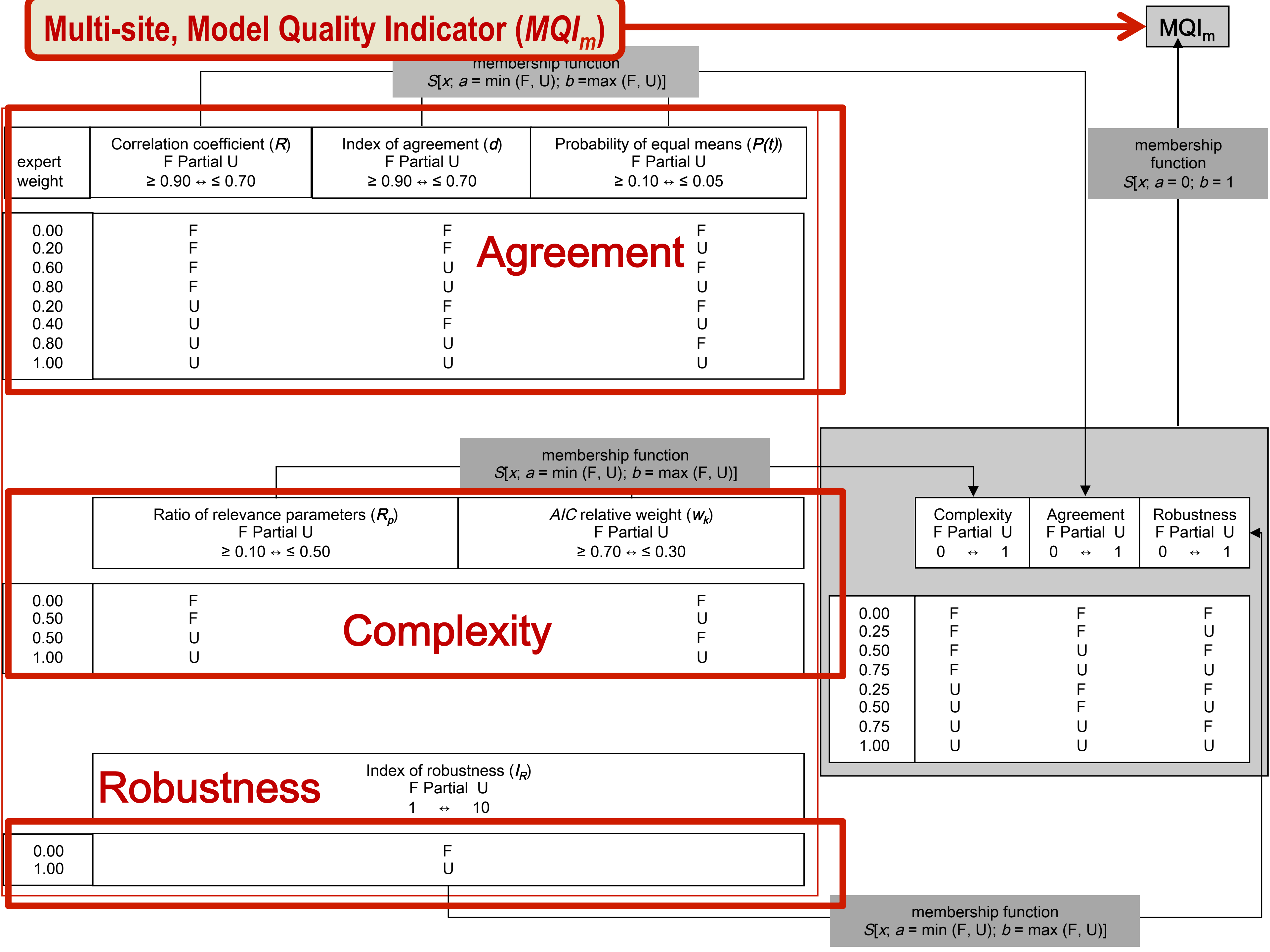
II. Complexity
- Ratio of relevant parameters
- Parameters-agreement criterion

III. Stability (robustness)
- Index of robustness

Model Quality Indicator



Hindrances to overcome: thresholds and weights



Multi-site, Model Quality Indicator (MQI_m)
MQI_m
membership function
$S[x; a = \min(F, U); b = \max(F, U)]$
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
Agreement
membership function
$S[x; a = 0; b = 1]$
membership function
$S[x; a = \min(F, U); b = \max(F, U)]$
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
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
$S[x; a = \min(F, U); b = \max(F, U)]$

MQI$_m$ – Questionnaire

Questionnaires answered / commented: 16 (13 online + 3 offline) + 1 comment

7. Over the range 0 (best) to 1 (worst) of *MQI*$_m$, may **crisp threshold values** be set to interpret results (e.g. >0.66: poor model performance?

6. Do the expert **weights assigned to Modules** reflect the importance of each of them?

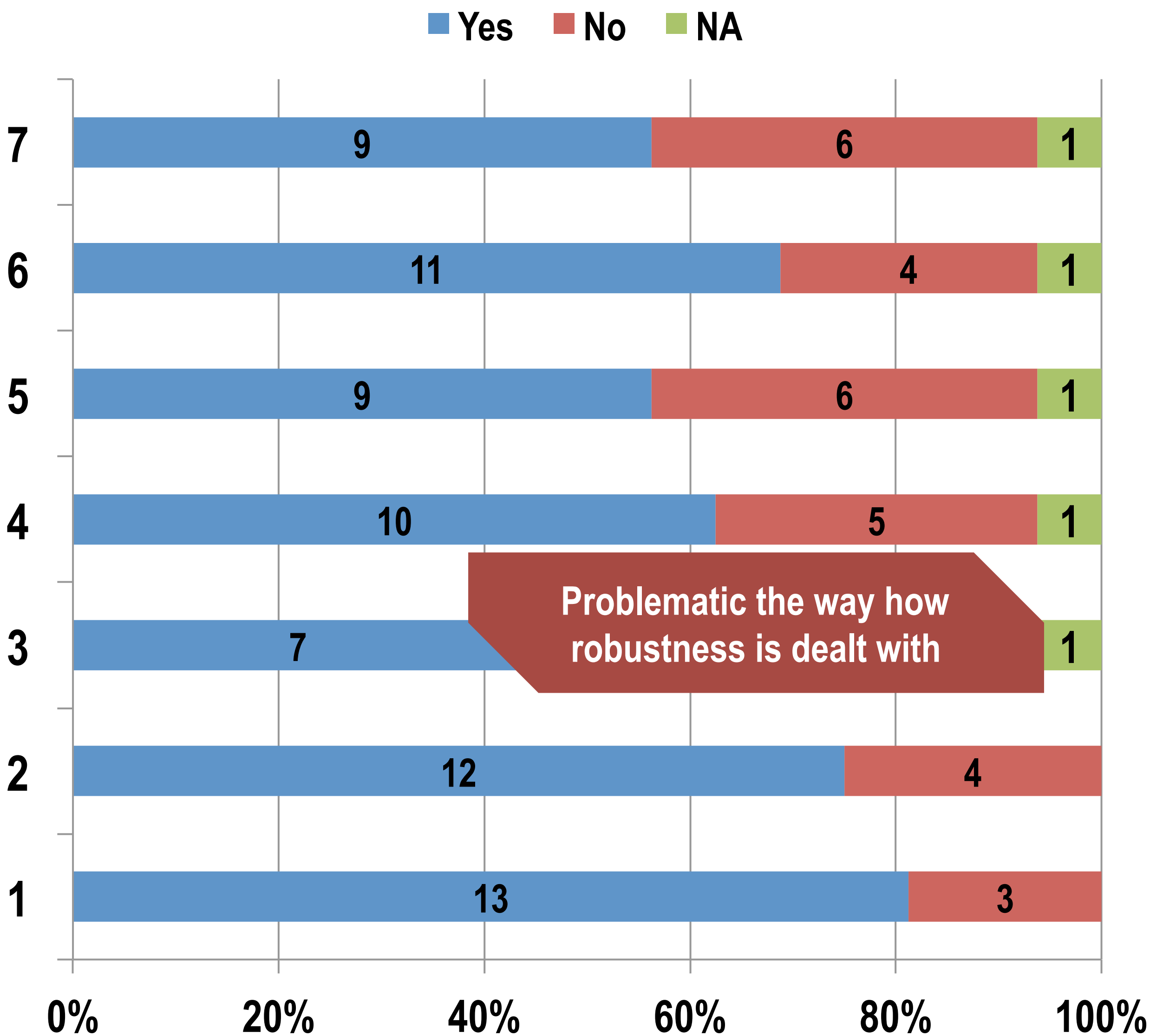
5. Do the expert **weights assigned to metrics** within a Module reflect their relative importance?

4. Do the **favourable / unfavourable thresholds** assigned to each metric reflect the perception of the quality of model performance?

3. Do the **equations** of the metrics need changes?

2. Do the metrics of *MQI*$_m$ represent a good choice to cover **aspects of model evaluation** (quantification of error, bias, efficiency, etc.)?

1. Do the fuzzy-logic based assessment method (*MQI*$_m$) account for all the **relevant aspects of model inter-comparison**?



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

Confalonieri et al. (2010)

$$I_R = \frac{\sigma_{EF}}{\sigma_{SAM}}$$

(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)

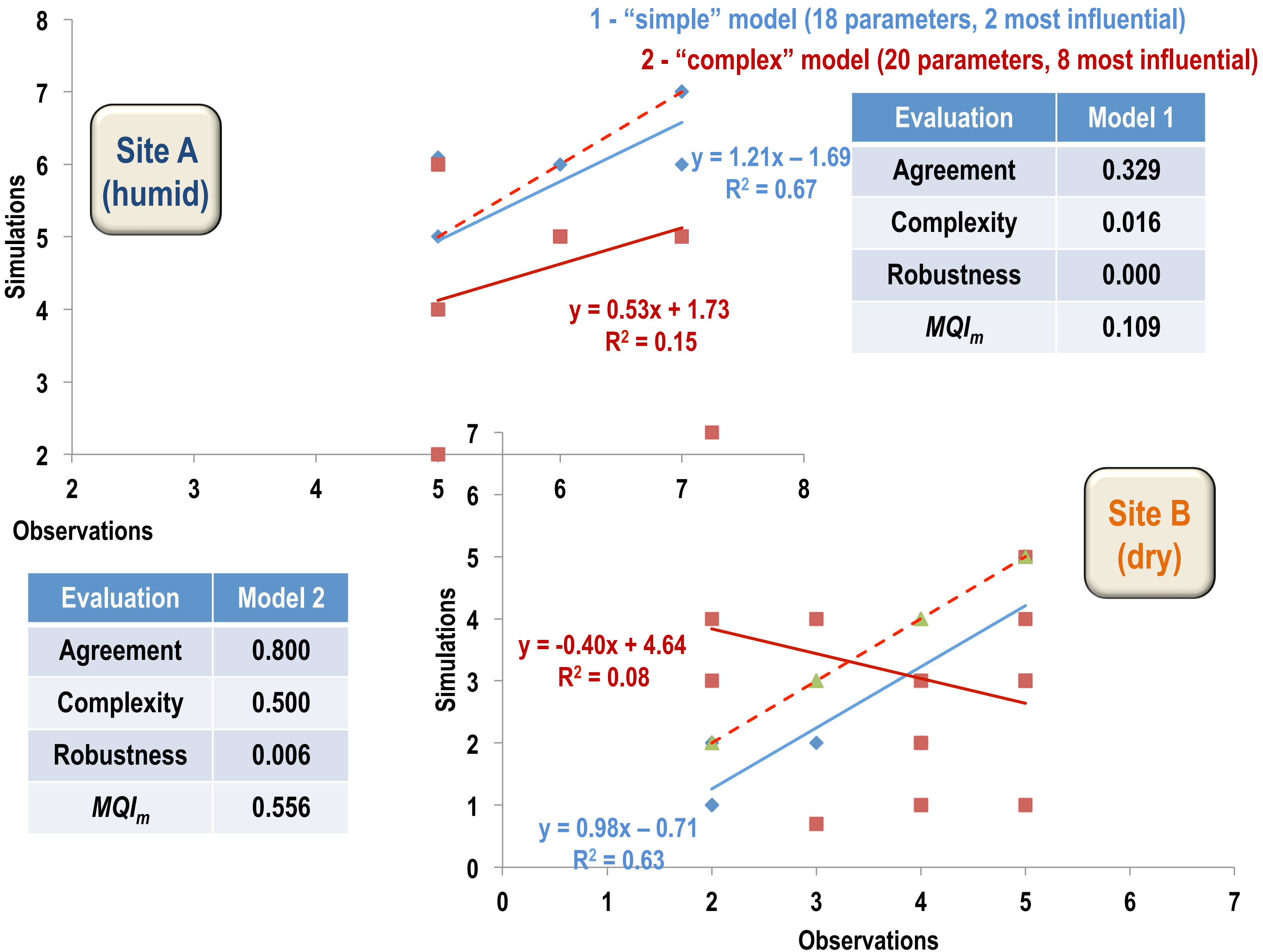
Synthetic Agro-Meteorological Indicator

$$SAM = \frac{Rain - ET_0}{Rain + ET_0}$$

(-1, +1)

From the questionnaires:

- Need to test the index on a variety of rainfall patterns (e.g. monsoonal areas)
- Whole year versus growing season, or winter and summer?
- Accounting for soil properties if water limited simulations are performed



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

Exemplary results

Above-ground rice biomass (kg DM m^{-2})

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

MQI_s	WARM	CropSyst	WOFOST
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	WARM	CropSyst	WOFOST
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

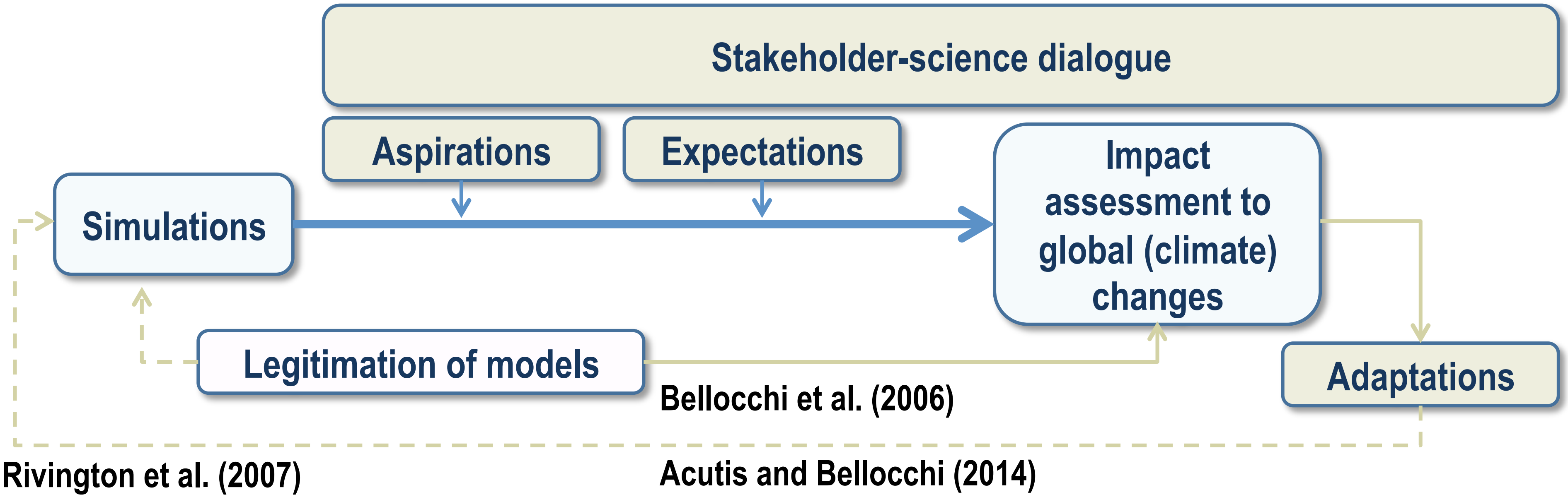
Robustness

MSE	WARM	CropSyst	WOFOST
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	WARM	CropSyst	WOFOST
C. d'Agogna	**34**	37	**79**
Vercelli	**33**	34	**73**
Mortara	**26**	28	**67**
Rosate	**20**	49	**91**

Complexity

Deliberative process in model-based climate change studies



Implementation and resources / 1

... | 2010 | 2011 | 2012 | 2013 | 2014 | 2015 | 2016 | 2017 | ...

AgMIP

JPI FACCE : MACSUR, CN-MIP, ...

MACSUR knowledge hub (as well as parallel programmes such as AgMIP or other initiatives of the JPI FACCE) holds potential to advance in good modelling practice in relation with model evaluation (including access to appropriate software tools), an activity which is frequently neglected in the context of time-limited projects.

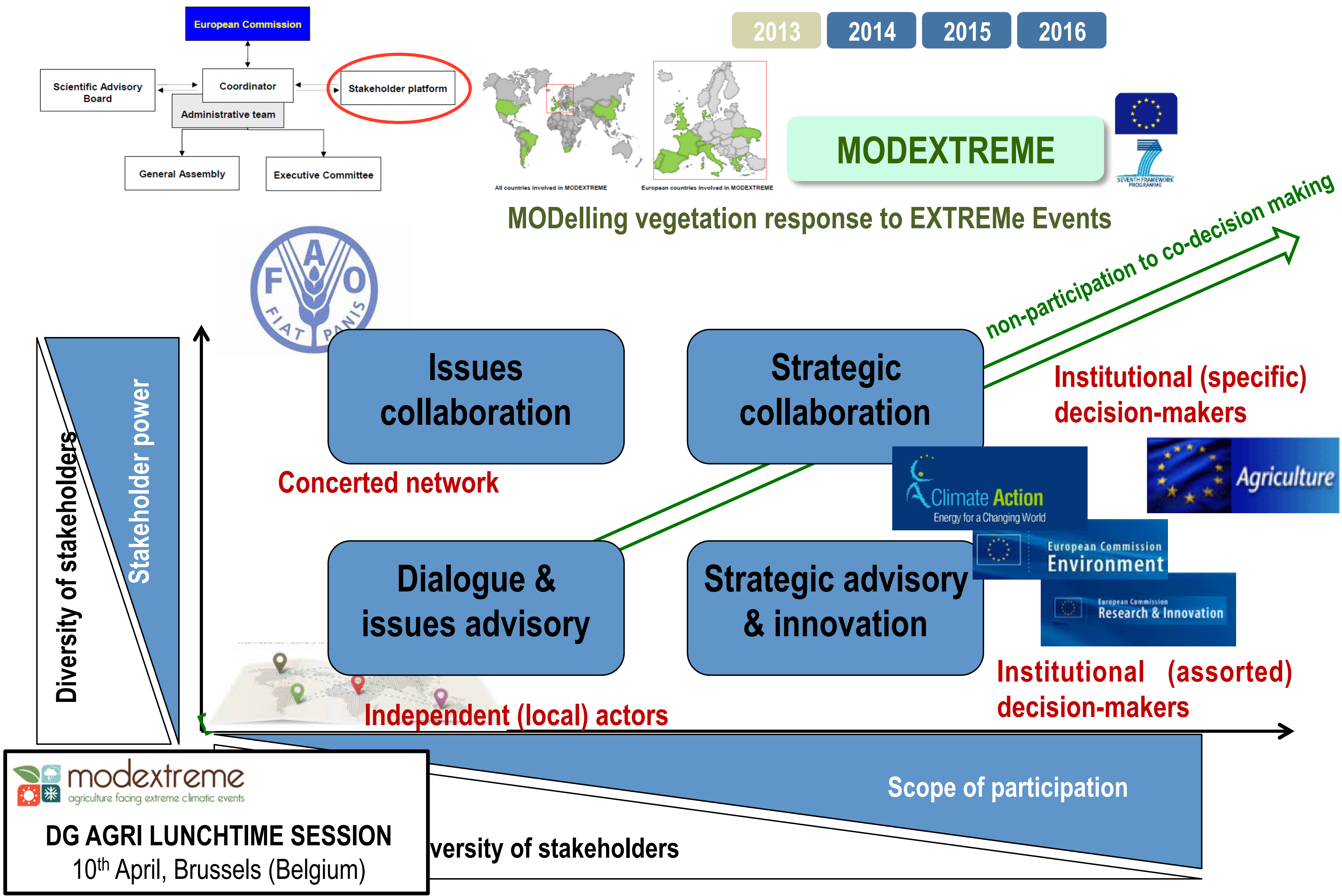
MACSUR

MACSUR Mid Term Conference
1st-4th April, Sassari (Italy)

LiveM →

International Livestock Modelling and Research Colloquium
14th-16th October, Bilbao (Spain)

Implementation and resources / 2



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 (crop and grassland) simulation models 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)

"*We conserve many things that we don't evaluate and little of those we value*" (Geoffrey M. Heal)