FACCE MACSUR

Model comparison and improvement: Links established with other consortia

Report on Task H1-XC1 - Sub-task XC1.3

Gianni Bellocchi[1]*, Lorenzo Brilli[2], Roberto Ferrise[3], Camilla Dibari[3], Marco Bindi[3]

[1] UMR Ecosystème Prairial, INRA, 63000 Clermont-Ferrand, France
[2] CNR-Ibimet, Via Caproni 8, 50145 Florence, Italy
[3] University of Florence, P.le delle Cascine 18, 50144 Florence, Italy

*gianni.bellocchi@inra.fr

Instrument:	Joint Programming Initiative
Topic:	Agriculture, Food Security, and Climate Change
Project:	Modelling European Agriculture with Climate Change for Food Security (FACCE-MACSUR)
Start date of project:	1 May 2015 (phase 2)
Duration:	24 months
Theme, Work Package:	WP H1 Cross-cutting activities
Deliverable reference num.:	D-XC1.3
Deliverable lead partner:	INRA
Due date of deliverable:	June 2016
Submission date:	2017–03–31
Confidential till:	

Revision	Changes	Date
1.0	First Release	2017-04-04

Instrument:	Joint Programming Initiative
Topic:	Agriculture, Food Security, and Climate Change
Project:	Modelling European Agriculture with Climate Change for Food Security (FACCE-MACSUR)
Start date of project:	1 May 2015 (phase 2)
Duration:	24 months
Theme, Work Package:	WP H1 Cross-cutting activities
Deliverable reference num.:	D-XC1.3
Deliverable lead partner:	INRA
Due date of deliverable:	June 2016
Submission date:	2017–03–31
Confidential till:	

Revision	Changes	Date
1.0	First Release	2017-04-04

Abstract

XC1 has established links to other research activities and consortia on model comparison and improvement. They include the global initiatives AgMIP (http://www.agmip.org) and GRA (http://www.globalresearchalliance.org), and the EU-FP7 project MODEXTREME (http://modextreme.org). These links have allowed sharing and communication of recent results and methods, and have created opportunities for future research calls.

1. Introduction

MACSUR-XC1.3 was meant to complement other projects and initiative in contributing toward integrating evaluation approaches in agricultural modelling in Europe and beyond Europe. Several research project and initiatives related to agricultural modelling are confronted with the need of ensuring quality in model performances. They often reflect the authors' perception of issues that are fundamental to understanding the factors that are related to model evaluation. They also demonstrate how previous instances of model use (and success or failure associated with that use) are the growing knowledge bases acquired from using different models for various applications. Scope and capabilities of evaluation approaches have evolved and have been improved with time. Though finding solution of how best to evaluate numerical values produced by models will remain an issue, a range of approaches do exist for improving the testing of model estimates. They include: disaggregation of evaluation statistics into basic components, introduction of evaluation criteria, and combination of statistics into synthetic indicators. These approaches continue to evolve. Baseline thresholds of evaluation measures provide users with the modellers' perception of good/bad performance statistics. Such criteria are presented and discussed not only to make available reference values of possible use in future evaluation studies, but also they call on the need for using expert rules to guide the evaluation process. Greater value can be gained through combined use and rule-based aggregation of multiple approaches to achieve a more complete form of evaluation.

Advancements in numerical testing methodologies for evaluation need to be put into structured frameworks comprised of processes such as sensitivity and uncertainty analyses (parameter and input variable appraisal), parameter optimization, model structure assessment (expert review), software testing, etc. In that perspective, evaluation is seen as an integral part of the overall model development and application process, whilst also encompassing the requirement for better data quality control and meta-data recording. This places a greater emphasis on the need to include evaluation plans within model project proposals and a higher level of support for evaluation work by funding organisations. Models may come in a variety of time and space resolutions and scales. Matching these scales and ensuring consistency in the overall model is not a trivial process and may be difficult to fully automate. Techniques to evaluate models need to be developed at the same pace with which the models themselves are created, improved and applied. Also, evaluation steps must be clearly stated, accessible, transparent, and understandable to non-modellers. In the context of the current knowledge, this can be achieved by means of reliability statistics, history of previous use, or personal preferences.

Model evaluation is a multifaceted complex process that is strongly influenced by the nature of the models as well as the conditions where they are applied. There is an increasing interest in the use of biophysical models to analyze agro-ecological systems, quantify outcomes, and drive decision making. Modeling applications have increased in the last decades, and the concept of model-based simulation of complex systems sounds attractive to support problem solving. However, problems exist when systematic and generalized evidence based on abstract knowledge is used by modelers, leaving potential model beneficiaries with less influence on decisions.

The participatory and deliberative feature suggests that the beneficiaries of model outputs may voice their complaints and desires to the model providers, discuss with each other and with the model providers, and, to some extent, influence and take responsibility for model content. A transition from model evaluation as academic research toward model evaluation as a participative, deliberative, and dialogue-based exercise (illustrated with two examples from international projects) is therefore desirable to raise the bar of model credibility and thus legitimate the use of agro-ecological models in decision making. Currently, the software technology to assist participatory approaches for model evaluation exists. The major limitation remains the difficulty to establish disciplined approached, effective management, and well-educated personnel within the time limitation and budgetary constraints of research projects. However, the continuing interest in the use of agro-ecological models to set ground for decisions offers opportunities to look at model evaluation with a fresh angle of vision and to question about opening new ways to see the principles of deliberative processes and software model development to converge.

Whether evaluation is a scheduled action in modelling, little work is published in the open literature (e.g., conference proceedings and journals) describing the evaluation experience accumulated by modeling teams (including interactions with the stakeholders). Failing to disseminate the evaluation experience may result in

the repetition of the same mistakes in future modelling projects. Learning from the past experience of others is an excellent and cost-effective educational tool.
The return on such an investment can easily be realized by preventing the failures of modelling projects and thus avoiding wrong simulation-based decisions.

The experience of MACSUR – Modelling European Agriculture with Climate Change for Food Security

The MACSUR (http://www.macsur.eu, 2002–2015) knowledge hub (as well as parallel programs such as AgMIP, http://www.agmip.org, or other initiatives of the FACCE JPI, http://www.jpifacce.org) holds potential to help advance good modeling 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. In MACSUR CropM-LiveM (crop-livestock modeling) cross-cutting activities, a questionnaire-based survey (through http://limesurvey.macsur.eu) on fuzzy logic-based multi-metric indicators for model evaluations helped understanding of the multifaceted knowledge and experience required and the substantial challenges posed by the deliberative process. A composite indicator, elaborated by a limited group of specialists, was first revised by a broader representative group of modelers and then assessed via questionnaire survey of all project partners (scientists and end users, including trade modelers), the results of which were presented in an international conference. The indicator aggregates the three components of model quality-agreement with actual data, complexity, and stability-represented. Seven questions were asked about indicator characteristics, which were answered by 16 respondents. The responses received reflect a general consensus on the key terms of the original proposal, although caution is advised on how metrics were formulated. In particular, some remarks and considerations suggested that other factors than purely climatic ones (such as soil conditions) may play a role in the concept of robustness and the construction of its metric.

Links established with other agricultural modelling projects and initiatives

MODEXTREME - MODelling vegetation response to EXTREMe Events

The EU-FP7 project MODEXTREME (http:///www.modextreme.org, 2013–2016) is an example of science-stakeholder dialogue where a platform of diverse stakeholders (from local actors to institutional parties) is established to evaluate crop/grassland/tree models. The aim is to represent the impact of extreme weather events on agricultural production. The strategy is to develop and implement modelling solutions and couple them with numerical evaluation tools, mostly using the capabilities of the platform BioMA (http://bioma.jrc.ec.europa.eu). The mass of stakeholder engagement reveals four clusters:

- The cluster "dialogue and issues advisory" demonstrates a high diversity of stakeholders with low power; i.e., broad types of stakeholders within operational and managerial scope (farmers, providers of agricultural services, field research agronomists) are identified locally, mainly by non-European project partners (Brazilian Corporation of Agricultural Research, Argentinian National Agricultural Technology Institute, University of Pretoria in South Africa, Chinese Academy of Agricultural Sciences).
- The cluster "issues of collaboration" is characterized by a partner (Food and Agriculture Organization of the United Nations) with considerable power, regarded as a stakeholder for the clear understanding of specific issues (food security) beyond the scope of the project and within a limited scope (local communities).
- In the cluster "strategic collaboration", stakeholders are a limited group of institutional actors (at the level of European Commission), regarded as partners for their direct involvement in research actions via survey techniques, meetings with representatives, and exchange of datasets (http://modextreme.org/event/dgagri2014). The power is high because the Joint Research Centre and the directorate for agriculture have the control to transfer scientific advances from the project into knowledge suitable for policy implementation in Europe (e.g., in-season crop monitoring and forecasts, integrated assessments in agriculture, and price regulation of agricultural commodities).
- The final cluster "strategic advisory and innovation" leads to institutional diversity, still at the level of European Commission. In contrast with the strategic collaboration predominant in the prior cluster, this cluster advises the dissemination strategy broadly (large scope extending to climate, environment, energy, and research), with less power for implementation.

The tool IMMA (Integrated Multi-metrics Model Analyzer) has been revised to statistically analyze simulation outputs (comparison of estimated and measured series of data) as part of the suite of functionalities of the modelling platform BioMA. Several trainings, meetings and workshops took place during the project's time life to address specific issues facilitating the model evaluation task and comparison between the alternative modelling solutions. During stakeholders' meetings project partners presented their results associated with the evaluation of modelling solutions, and received feedback. Model evaluation was performed against experimental data representing both usual and extreme weather conditions, using site-specific field data and observational sites as well as gridded data.

Model evaluation against observed data highlighted some limitations in the response to extreme weather events of state-of-the art modelling solutions, as emerged from the assessment performed at a variety of sites and systems worldwide. At the same time, the improvement introduced by modelling solutions modified was documented, as a basis for a discussion around the employment of modelling tools to support food security issues in Europe and worldwide.

However, it is difficult to draw general conclusions when the agro-ecosystems assessed are diverse and the exercise is based on limited weather and plant data.

Model Intercomparison for agricultural GHG emissions

A modelling study supported by five research projects (CN-MIP, Models4Pastures, MACSUR, COMET-Global and MAGGNET), funded by a multi-stakeholder call on agricultural GHGs with the support of FACCE JPI, and coordinated by the Integrative Research Group (IRG) of the Global Research Alliance (GRA), assessed the uncertainties in crop and grassland ensemble model simulations of productivity and N_2O emissions. This study results from a major crop and grassland model comparison and benchmarking exercise (Soussana et al., 2016; Sandor et al., 2016). It has highlighted the challenges faced in the implementation of a multi-model ensemble for forecasting agricultural productivity, nitrous oxide (N_2O) emissions and soil carbon stock changes. A total of 24 process-based biogeochemical models were assessed against long-term experimental field data from five grassland sites and five arable crop rotation sites from four continents. Model evaluation included uncalibrated (blind), partially- and fully- calibrated simulations through five modelling stages performed with sequential introduction of more detail of site-specific information. The accuracy of model estimates for above-ground productivity and N_2O emissions was evaluated at each stage by comparison with the experimental measurements and their related uncertainties. The comparison focused on agricultural productivity (grain yields for crops, and above-ground net primary productivity for grasslands) and annual N_2O emissions. It showed that the medians of the model ensemble predicted results more accurately and more consistently than any single model. From modelling evaluation stages 1 to 5, the relative root mean square error declined markedly for grain yields, and to a lesser and more variable extent for N_2O emissions and grassland above-ground net primary production (ANPP). The model ensemble median and a fraction of the individual models were within one standard deviation of the observed mean for N_2O emissions at most sites. Grain yield estimates were not within one standard deviation of observations until observed phenology (anthesis and maturity dates) was provided for calibration, together with productivity data. Reduced-size ensembles of 3-6 simulation models were shown to provide estimates that were within one standard deviation of the observed mean of N_2O emissions. The use of model ensembles as well as subsets of ensembles to estimate agricultural productivity and N_2O emissions at both field and regional scales have been discussed as a means to reduce simulation uncertainties.

AgMIP - The Agricultural Model Intercomparison and Improvement Project Project

AgMIP (https://www.agmip.org) is a major international collaborative effort to improve the state of agricultural simulation and to understand climate impacts on the agricultural sector at global and regional scales (Rosenzweig et al., 2013). To this, AgMIP brings together the climate, crop, livestock, and economics modelling communities and nutrition scientists to conduct improved integrated assessments of stresses facing the agricultural sector and interventions to overcome them. Among the main objectives of the project, improving agricultural models as well as exploring and managing uncertainties related to their application for impact assessment studies have a crucial role.

The evaluation, intercomparison, and improvement of crop models is organized by crop species. For each crop, multiple modelling groups used observed data from multiple sites for running the models. The range of model results as well as the discrepancies between simulated and observed values were analyzed. The former provides an indication of model variability, the latter offers an indication of model reliability (Wallach, 2006). Results from these inter-comparisons (e.g. Asseng et al., 2013 and Martre et al., 2015 for wheat; Bassu et al., 2014 for maize; Li et al., 2015 for rice; Fleisher et al., 2016 for potato) revealed a substantial variability between models that can be ascribed to uncertainties in both model structure and parameterization. Nonetheless, it was found that both the mean and median of a multi-model crop ensemble are better predictors than even the best individual model, and that the median is slightly better than the mean. Further, quantity and quality of observed data for model calibration had an important role in reducing uncertainties. Moving from low (providing only few data for calibrating models) to full (providing the whole set of observed data) calibration reduced the variation among the ensemble of contributing models as well as improved the accuracy of the mean and median of the ensemble.

AgMIP has contributed to enhancing the capacity for the agricultural climate change research community to conduct such model intercomparisons and improvements. Its activity is continuing by extending the analysis not only to other crop models but also to livestock and economic models.

3. Concluding remarks

Links established with other projects and initiatives in agricultural modelling helped to strengthen cross-learning about model evaluation. In particular, the interactions developed between a knowledge hub (MACSUR), a collaborative EU-FP7 project (MODEXTREME) and international initiatives (AgMIP and those supported by JPI FACCE projects) are a nice illustration of projects mutually building upon each other's community contacts and opportunities. The variety of experiences illustrated above indicate that there is still a difficulty to transfer brainstorming items into a formal model evaluation process based on agreed rules. In that, there are still missing opportunities in linking more closely the modelling developments and applications, stakeholder involvement and simulation-based decision making despite natural complementarities. However, these experiences confirm the role played by scientific knowledge hubs like MACSUR, which appear as critical pillars to advance good modeling practice in relation to model evaluation, an activity which is frequently neglected in the context of time-limited framework programs.

References

Asseng, S., Ewert, F., Rosenzweig, C., Jones, J.W., Hatfield, J.L., Ruane, A.C., Boote, K.J., Thorburn, P.J., Rötter, R.P., Cammarano, D., Brisson, N., Basso, B., Martre, P., Aggarwal, P.K., Angulo, C., Bertuzzi, P., Biernath, C., Challinor, A.J., Doltra, J., Gayler, S., Goldberg, R., Grant, R., Heng, L., Hooker, J., Hunt, L.A., Ingwersen, J., Izaurralde, R.C., Kersebaum, K.C., Müller, C., Naresh Kumar, S., Nendel, C., O'Leary, G., Olesen, J.E., Osborne, T.M., Palosuo, T., Priesack, E., Ripoche, D., Semenov, M.A., Shcherbak, I., Steduto, P., Stöckle, C., Stratonovitch, P., Streck, T., Supit, I., Tao, F., Travasso, M., Waha, K., Wallach, D., White, J.W., Williams, J.R., Wolf, J., 2013. Uncertainty in simulating wheat yields under climate change. Nature Climate Change 3, 827–832. doi:10.1038/nclimate1916

Bassu, S., Brisson, N., Durand, J.-L., Boote, K., Lizaso, J., Jones, J.W., Rosenzweig, C., Ruane, A.C., Adam, M., Baron, C., Basso, B., Biernath, C., Boogaard, H., Conijn, S., Corbeels, M., Deryng, D., De Sanctis, G., Gayler, S., Grassini, P., Hatfield, J., Hoek, S., Izaurralde, C., Jongschaap, R., Kemanian, A.R., Kersebaum, K.C., Kim, S.-H., Kumar, N.S., Makowski, D., Müller, C., Nendel, C., Priesack, E., Pravia, M.V., Sau, F., Shcherbak, I., Tao, F., Teixeira, E., Timlin, D., Waha, K., 2014. How do various maize crop models vary in their responses to climate change factors? Global Change Biology 20, 2301–2320. doi:10.1111/gcb.12520

Fleisher, D.H., Condori, B., Quiroz, R., Alva, A., Asseng, S., Barreda, C., Bindi, M., Boote, K.J., Ferrise, R., Franke, A.C., Govindakrishnan, P.M., Harahagazwe, D., Hoogenboom, G., Kumar, S.N., Merante, P., Nendel, C., Olesen, J.E., Parker, P.S., Raes, D., Raymundo, R., Ruane, A.C., Stöckle, C., Supit, I., Vanuytrecht, E., Wolf, J., Woli, P., 2016. A Potato Model Inter-comparison Across Varying Climates and Productivity Levels. Global Change Biology. doi:10.1111/gcb.13411

Li, T., Hasegawa, T., Yin, X., Zhu, Y., Boote, K., Adam, M., Bregaglio, S., Buis, S., Confalonieri, R., Fumoto, T., Gaydon, D., Marcaida, M., Nakagawa, H., Oriol, P., Ruane, A.C., Ruget, F., Singh, B.-, Singh, U., Tang, L., Tao, F., Wilkens, P., Yoshida, H., Zhang, Z., Bouman, B., 2015. Uncertainties in predicting rice yield by current crop models under a wide range of climatic conditions. Global Change Biology 21, 1328–1341. doi:10.1111/gcb.12758

Martre, P., Wallach, D., Asseng, S., Ewert, F., Jones, J.W., Rötter, R.P., Boote, K.J., Ruane, A.C., Thorburn, P.J., Cammarano, D., Hatfield, J.L., Rosenzweig, C., Aggarwal, P.K., Angulo, C., Basso, B., Bertuzzi, P., Biernath, C., Brisson, N., Challinor, A.J., Doltra, J., Gayler, S., Goldberg, R., Grant, R.F., Heng, L., Hooker, J., Hunt, L.A., Ingwersen, J., Izaurralde, R.C., Kersebaum, K.C., Müller, C., Kumar, S.N., Nendel, C., O'leary, G., Olesen, J.E., Osborne, T.M., Palosuo, T., Priesack, E., Ripoche, D., Semenov, M.A., Shcherbak, I., Steduto, P., Stöckle, C.O., Stratonovitch, P., Streck, T., Supit, I., Tao, F., Travasso, M., Waha, K., White, J.W., Wolf, J., 2015. Multimodel ensembles of wheat growth: many models are better than one. Global Change Biology 21, 911-925. doi:10.1111/gcb.12768

Rosenzweig, C., Jones, J.W., Hatfield, J.L., Ruane, A.C., Boote, K.J., Thorburn, P., Antle, J.M., Nelson, G.C., Porter, C., Janssen, S., others, 2013. The agricultural model intercomparison and improvement project (AgMIP): protocols and pilot studies. Agricultural and Forest Meteorology 170, 166–182. doi:10.1016/j.agrformet.2012.09.011

Sándor, R., Ehrhardt, F., Basso, B., Bellocchi, G., Bhatia, A., Brilli, L., De Antoni Migliorati, M., Doltra, J., Dorich, C., Doro, L., Fitton, N., Giacomini, S.J., Grace, P., Grant, B., Harrison, M.T., Jones, S., Kirschbaum, M.U.F., Klumpp, K., Laville, P., Léonard, J., Liebig, M., Lieffering, M., Martin, R., McAuliffe, R., Meier, E., Merbold, L., Moore, A., Myrgiotis, V., Newton, P., Pattey, E., Recous, S., Rolinski, S., Sharp, J., Massad, R.S., Smith, P., Smith, W., Snow, V., Wu, L., Zhang, Q., Soussana, J.F., 2016. C and N models Intercomparison – benchmark and ensemble model estimates for grassland production. Advances in Animal Biosciences 7, 245–247. doi:10.1017/S2040470016000297

Soussana, J.F., Ehrhardt, F., Snow, V., Sándor, R., Bellocchi, G., Grace, P., Recous S., 2016. Assessing simulation models for field scale projections of pasture and crop GHG emissions. 6th Greenhouse Gas and Animal Agriculture Conference, 14–18 February, Melbourne, VIC, Australia.

Wallach, D., 2006. Evaluating crop models. In: Wallach, D., Makowski, D., Jones, J.W. (Eds.), Working with dynamic crop models, Elsevier, Amsterdam, The Netherlands, pp. 11–53.