Towards systematic evaluation of crop model outputs for global land-use models

D. Leclère [1] – L. B. Azevedo [1] – R. Skalský [1]– J. Balkovič [1] – P. Havlík [1]
[1] Ecosystem Services Management, IIASA, Laxenburg (Austria), leclere@iiasa.ac.at

Abstract

Land provides vital socioeconomic resources to the society, however at the cost of large environmental degradations. Global integrated models combining high resolution global gridded crop models (GGCMs) and global economical models (GEMs) are increasingly being used to inform sustainable solution for agricultural land-use. However, little effort has yet been done to evaluate and compare the accuracy of GGCM outputs (Mueller & Robertson, 2014). In addition, gridded global crop model datasets require a large amount of parameters whose variability across space is poorly documented: increasing the accuracy of such dataset has a very high computing cost. Innovative evaluation methods are required both to ground credibility to the global integrated models, and to allow efficient parameterization of GGCMs.

We propose an evaluation strategy for global gridded crop model datasets in the perspective of use in global economic land-use models, illustrated with preliminary results from a novel dataset (the *Hypercube*) generated by the EPIC GGCM (zaurralde et al., 2006) and used to inform the GLOBIOM land use model (Havlík et al., 2014).

Material and Methods

Evaluation rationale & strategy. We apply to the Hypercube dataset the principles of model evaluation developed by (Jakeman et al., 2006): evaluation should provides a transparent diagnosis of model adequacy for its intended use. We below describe how the Hypercube data is generated and how it articulates with GLOBIOM (Figure 1), in order to transparently identify the performances to be evaluated, as well as the main assumptions and data processing involved.

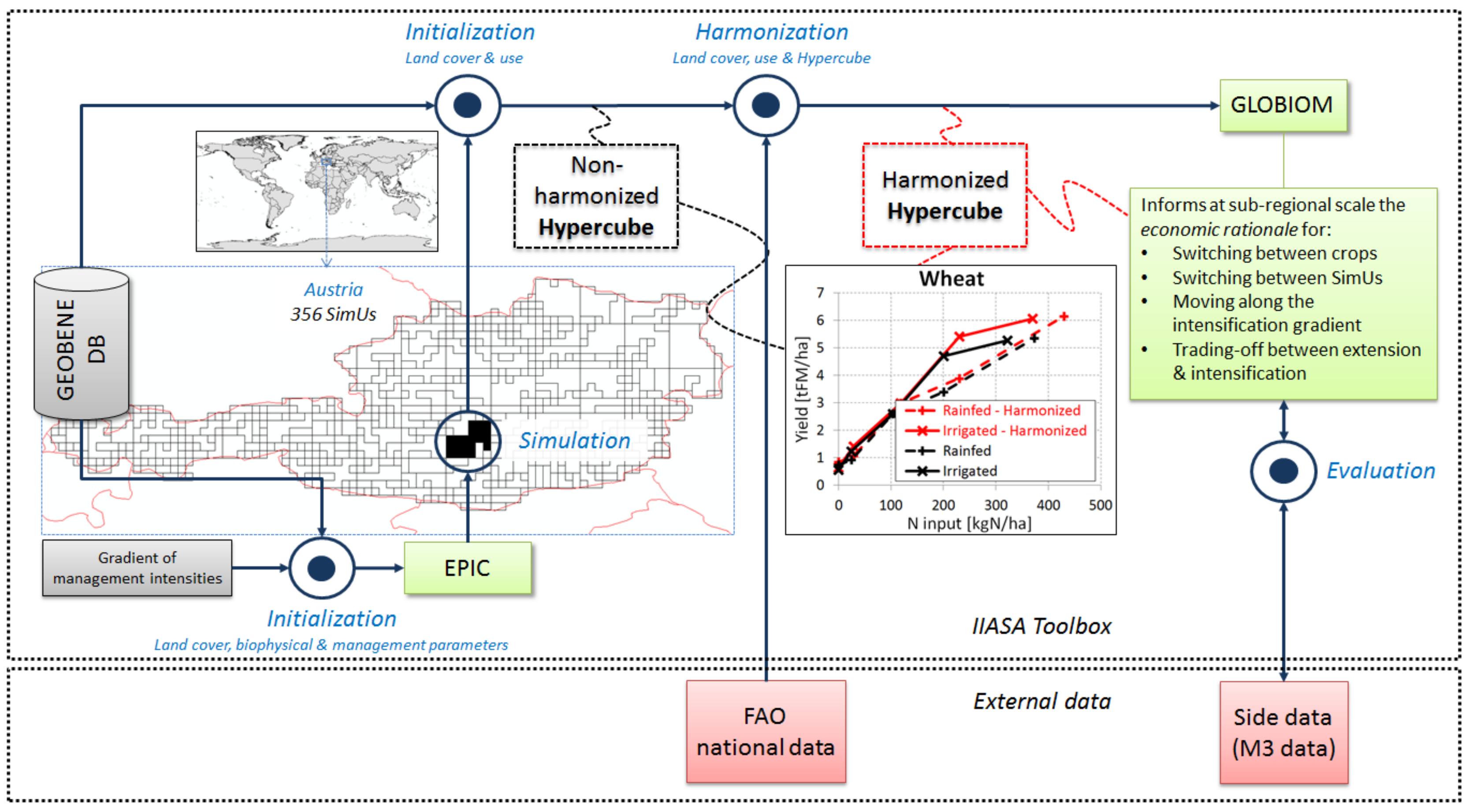

Figure 1. *Overview of the Hypercube dataset, its articulation to the EPIC and GLOBIOM models, and the information required to propose an evaluation strategy: data processing steps (initialization, simulation & harmonization) and use in the GLOBIOM model.*

The Hypercube dataset. The *GEOBENE* database for bio-physical modeling (Skalský et al., 2008) splits global land surface at 5 arcmin resolution into 212707 spatial units – SimUs – of homogeneous biophysical environment (present climate, soil, altitude and slope). It provides a description of land cover and use common to the *GLOBIOM* and the *EPIC* models. For each SimU identified as potential cropland, we feed these biophysical inputs into the EPIC model (v. 0810) to estimate for 16 major crops the yield, nutrient and water inputs around year 2000 (Balkovič et al., 2014). Simulations are done for a gradient of management varying by intensity level (5 N-application rates x 3 irrigation rate): overall the *Hypercube dataset* provides at high spatial resolution an estimate of the response of crop yield and input needs to a gradient of crop management practices intensification over potential cropland.

Use in GLOBIOM, initialization and harmonization steps. The Hypercube dataset is used in GLOBIOM to inform on the current and potential crop yields and inputs across simulation units and species along an intensification gradient. As described in Figure 1, we initialize the geographical land use information of crop and management practices acreage around year 2000 by overlaying the hypercube data with the GEO-BENE global database. Moreover, as GLOBIOM needs to provide estimates consistent with FAO data, an additional harmonization step scales the crop acreage and yield values to fit FAO national area, production and yield estimates. Finally, the harmonized dataset is used into GLOBIOM to characterize at high spatial resolution the economic rationale to alter the land use extent (acreage change by crop and across crops) and intensity (crop-specific management intensity).

REFERENCES

Balkovič et al., 2014, Glob. Planet. Change, 122, 107–121.; Havlík et al., 2014, PNAS, 111, 3709–; 14.; Izaurralde et al., 2006, Ecol. Modell., 192, 362–384.; Jakeman et al., 2006, Environ. Model. Softw., 21, 602–614.; Monfreda et al., 2008, Global Biogeochem. Cycles, 22, 1–19.; Mueller & Robertson, 2014, Agric. Econ., 45, 37–50.; Mueller et al., 2012, Nature, 490, 254–7.; Skalský et al., 2008, GEO-BENE global database for bio-physical modeling v. 1.0 (Concepts, methodologie and data). Laxenburg, Austria.

Preliminary results

Identified performances to be evaluated. We expect the *Hypercube dataset* to adequately represent sub-national heterogeneities in crop yield and input needs: *i*) in space, *ii*) across crop species, and *iii*) across management intensities. In addition, the sensitivity of these performances to the initialization step needs to be evaluated. Finally, the harmonization to FAO data should be factored in the evaluation, which should differentiate between the national and sub-national scales, and before and after harmonization.

National scale fit to FAO & sensitivity to initialization. After land use and management initialization and before harmonization, we compare the nationally aggregated Hypercube dataset to FAO yield estimates around year 2000 for the 121 countries considered in the GLOBIOM model. Table 1 provides the world averaged FAO reported yield (World avg.), as well as the root mean square error (RMSE), Nash–Sutcliffe model efficiency (NSE) and Spearman rank correlation (spear. cor.) coefficients for five crops and two different initializations (SPAM datasets v0 vs. v3). The performances vary by crop: the correlation is larger (resp. lower) than 0.5 for Maize, Barley, and Wheat (resp. Rice and Millet). The NSE is at best positive, and the RMSE is always larger than 50% of the world average yield. The influence of the initialization step (v0 vs. v3) seems limited, suggesting this noise stems mainly from assumptions on the EPIC parameters as well as general on the EPIC performance.

Crop	**Corn**		**Barley**		**Rice**		**Wheat**		**Millet**	
SPAM v.	**v0**	**v3**	**v0**	**v3**	**v0**	**v3**	**v0**	**v3**	**v0**	**v3**
World avg. [tFM/ha]	4.39		2.38		3.87		2.82		0.71	
RMSE [tFM/ha]	2.6	2.58	1.32	1.32	2.45	2.49	1.69	1.77	0.91	0.8
NSE [-]	-0.21	-0.22	0.14	0.12	-0.11	-0.19	0.07	0.04	-0.21	-0.24
spear cor. [-]	0.56	0.52	0.74	0.75	0.15	0.11	0.7	0.63	0.38	0.33

Table 1. *Comparison of preliminary non-harmonized Hypercube data to FAO reported yields at country level for the year 2000, using two versions of SPAM data (v0 & v3) to test for the impact of various land use initializations.*

Sub-national scale actual & potential yields compared to the M3 dataset. Figure 2 illustrates the evaluation of the expected performances of the harmonized Hypercube dataset at sub-national scale. It compares for maize in the USA around year 2000 the minimum, maximum and actual (with management intensity initialization) yields of the harmonized Hypercube dataset to the M3 dataset of reported yields (Monfreda et al. 2008). The spatial variability in actual yields is less contrasted in the preliminary Hypercube dataset than in the M3 dataset, and yields are underestimated in the Western part of the USA. While it questions assumptions on the EPIC model inputs and performance, potential yields are only slightly too lower in this region, and underestimated management intensity could largely contribute to the low actual yields.

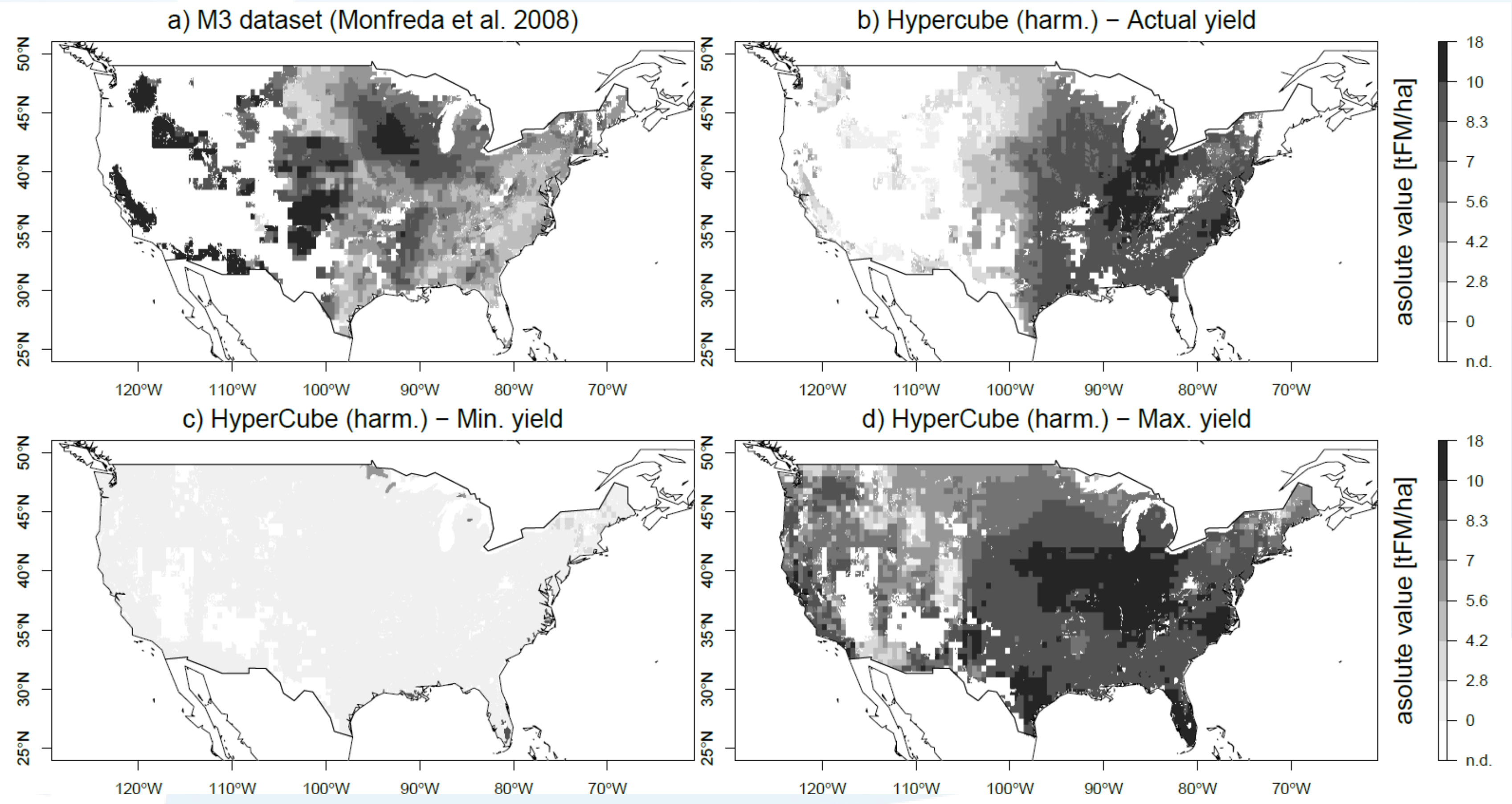

Figure 2. *High-resolution comparison for maize in the USA of the Hypercube data (preliminary results) and the M3 dataset: a) yields reported by Monfreda et al. (2008), and actual (b), minimum (c) and maximum (d) achievable yields in the harmonized Hypercube dataset.*

Take home message

- Crop model outputs used in integrated models lack evaluation: we propose an evaluation method illustrated with preliminary results from the EPIC-GLOBIOM IIASA model cluster.
- The evaluation exposes and measures the expected performances, and weights the relative contribution of crop model, input data and data processing steps.
- Such an approach targets future efforts for accuracy improvements and would achieve highest efficiency if combined with traditional field-scale model evaluation.
- Next steps include evaluation of current yield gaps and main limiting factors against the M3 dataset, iterative improvement of parameter assumptions, and definition of acceptable performances for intended use in the EPIC-GLOBIOM IIASA model cluster.