Geosci. Model Dev. Discuss.,
https://doi.org/10.5194/gmd-2018-295-AC2, 2019

Geoscientific Model Development Discussions
Open Access

Interactive comment on "MAgPIE 4 – A modular open source framework for modeling global land-systems" *by* Jan Philipp Dietrich et al.

Jan Philipp Dietrich et al.

dietrich@pik-potsdam.de

Received and published: 20 February 2019

We would like to thank the Referee for all the efforts and the valuable remarks which helped to considerably improve the paper.

[REFEREE COMMENT 1]: The paper presents the IT architecture of the MAgPIE framework focusing on two features: modularity and flexibility of spatial resolution. This is a rather technical paper which is well written and easily understandable in spite of its technicity. This type of paper is welcomed to improve the transparency of models and help interpreting their result. Here are my comments: 1. The presentation of the modules (p. 5-7) raise a number of issues: (i) The definition of the modules is sometimes vague. The "costs" module is not

Printer-friendly version

Discussion paper

easy to grasp: what kind of aggregates does it make? In fact, we wonder why this is a separate module for it. Why is the aggregation not done in the corresponding module? The "production" module is defined as aggregating cellular production to the regional level, but how the cellular production is defined? I would say this result from the "yields" and "crop" module, but this is not clear from the text and from Figure 1.

[AUTHORS RESPONSE 1]: In the revised manuscript we try to be more precise. In the case of the “costs” module we talk now about “summation” instead of “aggregation” specifically stating that the outcome is total costs to make it more transparent that it is not more than a total sum of costs. We changed the formulation to

[TEXT EDIT 1] “Calculates total costs by summing up all costs in the model including production costs, investments into research and development or land expansion, tax expenditures, and mitigation costs.”

[AUTHORS RESPONSE 1]: In case of module “production” we realized that the previous formulation was suggesting that crop production is calculated in the production module, which would actually be in contradiction to the visualization in Figure 1. In fact, crop production is already calculated in the crop module and just merged with other production information in the production module. To make this transparent we updated the descriptions of modules “production” and “crop” accordingly. We also added a sentence explaining for what purpose production values are aggregated to regional production in the production module. More specific questions such as the question how cellular crop production is defined are answered in the model documentation in the supplementary material and therefore not discussed here. New description of the “production” module is now:

[TEXT EDIT 1] “Merges production values including crop-based production and livestock-based production into one production variable. Aggregates cellular production to the regional level for modules only interested in regional production levels.”

New description of the "crop" module is now:

[TEXT EDIT 1] "Simulates crop production and competition of different crop types for cropland, accounting also for crop rotation requirements. Estimates the terrestrial carbon pools of croplands."

The updated formulation is now in line with the implementation in the model as well as the visualization in figure 1.

[REFEREE COMMENT 2]: (ii) Prices are almost absent from the picture while they are a key element of the system. They are the primary drivers of the intensification mechanisms which are for this reason unclear here: is there some livestock intensification? Does the technological change react to price or it is exogenous? Also how is the fertilizer use treated? In so doing we don't see the substitution possibilities between production factor while this is basically what the model represent.

[AUTHORS RESPONSE 2]: In MAgPIE prices are implicitly modelled as marginals of the model constraints and are therefore not part of the module descriptions. Intensification as well as all other decisions in the module are coming from an interplay of physical constraints and costs associated to activities in the model. We explain this now in the paper with the following sentence added to the brief history of MAgPIE section:

[TEXT EDIT 2] "Prices are implicitly modeled as marginals of the model constraints. Intensification as well as other decisions in the model are coming from an interplay of physical constraints and costs associated to activities in the model."

[AUTHORS RESPONSE 2]: The implemented intensification processes are explained in the model documentation. As explained in the description of the tc module, the available TC implementation is endogenous and intensification can be triggered via investments into technological change. Pasture yield intensification in MAgPIE 4.0 is available in 2 realizations (see yields module): In the realization "biocorrect" pasture yields grow parallel to crop yields based on TC investments, in “dynamic_aug18” pasture intensification is based on exogenous scenario assumptions. Besides pasture intensification there is also a scenario based conversion of livestock production systems implemented which is explained in detail in the model description of the livestock module. As many aspects are answered in the model documentation we make sure that it is more noticeable in the paper. Especially, we added a link to the documentation in the description of Figure 1 (model linkages).

[REFEREE COMMENT 3]: (iii) Finally some feedback loops seem to be lacking, e.g.: the production of residues should affect the bioenergy module; the crop module should affect the livestock through feed production; the livestock production may affect the yield through manure and the availability of land may have an impact on yields.

[AUTHORS RESPONSE 3]: A feedback of residue production on bioenergy demand is indeed not considered in the current setup. Instead the framework currently works with external demand scenarios for residues which are chosen consistently to the demand calculations in the bioenergy module. We are currently working on improving the residue implementation and expect changes for the next MAgPIE releases. The connection between livestock and crop production actually exists in two ways which are implied in Figure 1. The first connection goes via feed demand which triggers additional crop production (Livestock -> Demand -> Trade -> Production -> Crop). The second one goes through livestock triggering pasture demand which is competing with the crop module for land (Livestock -> Pasture -> Land -> Crop). Yield increases through manure are not explicitly modeled but land availability is affecting yields as expected by the reviewer. All these details about the interactions are explained in the model documentation and therefore not part of the paper. As mentioned above, we link the model documentation more noticeable in the paper. Furthermore, we explain now in the text at the end of subsection “Modules” that also modules not directly linked are connected to each other via other modules:

[TEXT EDIT 3] "If modules are not directly linked it does not mean that they do not interact with each other. In some cases the feedback loops go through a combination of modules rather than being direct links. An example is the livestock module which is triggering feed demand in the demand module, which is, via trade and production module, triggering production in the crop module."

[REFEREE COMMENT 4]: The last two points reveal the difficulty of representing a system in a modular way, as each module strongly interacts with the other, making the frontier between them sometimes meaningless. Livestock and crop production system are typical examples as they are generally strongly integrated. This point is an important barrier to the modular representation which should be discuss in depth and better justified. In some cases the modular representation is appropriate for modeling reality, however in some cases it could put into question the consistency of the model. Most importantly, this approach may be seen as only compatible with conventional agriculture, and not with alternative agricultural systems promoting a systemic approach.

[AUTHORS RESPONSE 4]: This is a very important remark which requires some discussion. Our experience so far showed the general applicability and usefulness of a modular structure is also quite high for strongly integrated systems. One aspect which we avoided to discuss so far, but is relevant in this context is the question of persistence of modular structures. Having a completely static modular structure would indeed significantly limit the modeling capabilities of MAgPIE. What we are dealing with instead is a semi-static structure, in which module definitions are valid on a longer time scale than their underlying realizations, but are also allowed to change from time to time. Introducing or removing interfaces is possible as well as creating, splitting, merging or deleting of modules. Doing so allows to adapt the framework structure to new challenges which cannot be reflected in the given structure. Concerning module separations in closely linked systems our experience is that this is useful as well and helps to understand the interactions better. The difference between closely and loosely linked modules is usually just the number of interfaces and interactions. What is gained by the modular structure is a better control over interactions in the model as only specifically designed interactions are possible, while all other interactions will be detected as a code violations. In the early stages of modularizing MAgPIE this uncovered for instance cross-links in MAgPIE in which completely unrelated variables were used as surrogates for processes not covered by the model. While this might be desirable in some cases it can easily lead to unrealistic model behavior when users are not aware of such a link and by default do not expect that such a link exists. A modular structure does not generally prohibit such links but it makes them clearly visible and forces the researcher to think about it. We extended the main text by adding a discussion addressing this issue:

[TEXT EDIT 4] "One main improvement introduced in MAgPIE 4 is the full code modularization. It is used as a tool to make the model better manageable as it structures the code in self-containing components which are interacting via interfaces with each other. It makes existing and missing interactions in the model better visible and allows to easily replace components by alternative implementations. While the modular structure is rather intuitive for a system with loosely linked components one could argue that it might prevent a proper implementation of strongly integrated systems. Our experience is that, while the modular concept is working best for clearly separable systems, it also works in all other cases. The difference with strongly integrated systems is that the amount of interfaces and the required effort for developing new realizations are higher. Nevertheless, it still improves transparency in terms of model interactions and does not exclude any systems or dynamics from being represented in the model. Modules are also not static and the modular structure itself can and will also be changed if required. Modules might get created, deleted, merged or split over time. Module interfaces might get extended, reduced or modified. As both happens less frequently than changes within modules the modular structure can be best described as semi-static."

[REFEREE COMMENT 5]: 2. The key evaluation examples are not very informa-

tive in the context of this paper. The paper presents the "framework" and not really the model (i.e., the economic and biophysical mechanisms represented). For this reason, having, e.g., an evaluation of the crop yield simulated by the model is not very relevant to the paper. We would rather expect an evaluation of the modular architecture, as the authors did for the spatial resolution in the next section. How the modules perform running together vs standalone?

[AUTHORS RESPONSE 5]: We agree that the key evaluation examples are a bit unrelated to the topic of the paper. We like the suggestion to show module-related examples instead. Consequently, we moved the key evaluation example to the appendix and added a new section "Impact of module realizations" to the model outputs section. The new section discusses three different applications of the modular structure (two cases in which alternative realizations are used and one case in which a module is run standalone). The new text reads as follows:

[TEXT EDIT 5] "Figure 2 shows three different applications of the flexible, modular structure in MAgPIE in comparison to a run with default settings. The first application (soil organic matter) is a case in which a model feature can be either switched on or off. While this module is slightly improving the overall accuracy of the model through improved fertilizer estimates it has high computational requirements, nearly doubling the run-time of the model. By default it is switched off but can be activated when needed, e.g. for studies focusing on fertilizer application. The second application (volume-based factor costs) is an example of a dispute about the representation of a process, in this case the relationship between factor requirement costs and production. We compare here two realizations of factor requirement costs, one of which mainly links them to the area under production (default realization) and the other of which mainly links them to the production itself. As the available data sources did not allow to clearly link costs to area or production we were experimenting with different realizations of it. The flexible modular structure allowed to easily implement different hypotheses and compare them which each other. The third application (standalone

food demand) is an example in which a module is enabled to run standalone. Here, the food demand calculations, estimating regional food demand based on GDP projections and demographics, can also be run independent of other modules. This is especially useful for studies focusing on food demand itself or for general improvements in the projections itself.

The evaluation plots show different stages and major components of a MAgPIE simulation. As figure 2 shows the population, which is an exogenous parameter driving the simulations, is identical for all four runs. As one of the drivers of food demand, the population is also available in the food demand standalone case. We get a similar picture for the per capita food demand, which is the main output of the food demand model. The output is available for all runs and due to identical scenario assumptions also identical (for different assumptions see a variation across SSPs in Appendix A2).

As all other aspects shown in the figure go beyond what is used or simulated in the food demand module, all remaining values could only be reported by the non-standalone runs. The combination of per capita food demand and total population provides the total food demand in the model which triggers total feed demand through livestock consumption. Also here the identical scenario assumption leads to the same results in all three runs. Differences can be observed in the global land cover and the productivity measures (land use intensity and average crop yields). Cropland shows higher expansion in the alternative scenarios compared to the default scenario while both scenarios show less intensification and lower yields. While the differences are rather small in the case of soil organic matter, the difference are quite pronounced in the alternative factor requirement case. In the case of soil organic matter this effect is triggered via the natural availability of nitrogen in the soil. Having SOM switched off the model assumes, that all required nitrogen has be provided as fertilizer, while simulating SOM explicitly uncovers the already available nitrogen in the soil. This reduces the overall fertilizer requirements and slightly incentivizes land expansion as it gives the model access to more nitrogen. As the food demand is rather independent of this decision *more land expansion leads to lower intensification requirements, lowering land use intensity as well as average yields. Having factor requirements primarily linked to the production rather than to the area on which it is produced strongly reduces the incentive in the model to intensify. Area dependent factor requirements strongly favor high yielding locations for production giving the model a strong incentive to concentrate production on high productive areas and to further boost productivity via intensification. Production dependent factor requirements on the other hand do not favor locations based on productivity making also rather unproductive areas interesting for production and thereby reducing the incentive for intensification. In combination this leads to significantly higher cropland expansion, higher forest reduction, less intensification and significantly lower crop yields. One can also observe that the difference in average yields is higher than in land use intensity, owing average yields to drop for two reasons: the lower land use intensification and the expansion into low productive areas.*

CO2 emissions show strong fluctuations in all scenarios due to missing constraints linking carbon stocks with the goal function of the model (e.g. carbon pricing). This makes it in many cases an arbitrary decision for the optimizer to expand cropland into carbon rich or carbon poor areas. Besides its fluctuations the plot also shows higher overall emissions in the case of volume-based factor costs due to the overall higher expansion of cropland and reduction in forest areas."

[REFEREE COMMENT 6]: 3. The evaluation of the spatial flexibility is interesting however I did not really understand why there is so much difference in the default and Brazil-specific settings. Why is there much specialization in the default setting? And to what extent does it affect the spatial deforestation/reforestation pattern in Brazil? Also, the paragraph beginning on p13 l4 is quite difficult to understand (why is there 200 clusters in the default version and 500 in the region-specific one ?)

[AUTHORS RESPONSE 6]: We reformulated the paragraph previously beginning on p13 l4 to make it easier to understand and to better explain the increase of clusters from 200 to 500:

[TEXT EDIT 6] "Figure 4 shows a setup with a specific focus on Brazil. To gain higher spatial detail in Brazil it comes with a higher number of clusters in total. Brazil (BRA) is simulated as a world region together with its most important trade partners (Rest of Latin America (LAM), United States (USA), China (CHA) and Europe (EUR)). Remaining countries, less relevant for a Brazil-centric study, are merged to a single region (ROW). Furthermore, the cluster allocation of 500 clusters in total has been shifted in favor of Brazil: Roughly four times more clusters are allocated to Brazil (306) compared to a default distribution of clusters. At the same time the rest of the world region receives only roughly 0.7 times the number of clusters it would usually get (37), leaving room for a balanced number of clusters for all other regions. Detail gained for Brazil is attained with reduced detail for the rest of the world to keep the model complexity manageable for the applied solver."

[AUTHORS RESPONSE 6]: To explain the observed specialization in the spatial patterns we added another section to the discussion addressing this issue:

[TEXT EDIT 6] "The observed specialization is a consequence of the homogeneous biophysical characteristics within each cluster which lead to either-or-decisions in the model. It will either fully take a cluster into production or ignore it completely. In the default setup this effect is very pronounced due to the low number of clusters within Latin America. With more clusters, as in the Brazil setup, clusters better grasp the real spatial distributions of biophysical characteristics in the region and therefore lead to a more diverse picture. Whereas this effect is especially relevant for regional studies with focus on spatial patterns, it is less critical for global dynamics as long as the spatial aggregation is not introducing any systematic biases to the model."

[AUTHORS RESPONSE 6]: Furthermore, we improved the discussion of forest cover development in Latin America and globally under both setups:

[TEXT EDIT 6] "Comparison with historical data sets as well as projections on for-

est cover show that the differences between mappings are rather small compared to the overall uncertainty in these numbers. Nevertheless, a deeper look into the simulations uncovers that the global numbers of the Brazil-centric setup are unreliable as the reduced deforestation rate compared to the default setup is a consequence of the applied mapping. As the ROW region basically acts as a huge free trade region it can fulfill strong demand pressure coming from Sub-Saharan Africa with production from elsewhere, while trade limitations in the default setup limit this exchange and trigger deforestation within Sub-Saharan Africa Africa (Dietrich, 2018, compare m4p_default_validation.pdf p1558 and m4p_brazil_validation.pdf p1465). In the case of LAM both runs show a rather similar picture in the aggregated forest cover projections for the region and it is not possible to clearly reject one of them. This is particular important as the regional aggregates in LAM are in the scope of both mappings and therefore should be sound. When choosing between them, one has to decide whether spatial details in Brazil or global trade patterns are the more decisive factor for accurate estimates of regional forest cover in LAM."

Interactive comment on Geosci. Model Dev. Discuss., https://doi.org/10.5194/gmd-2018-295, 2018.

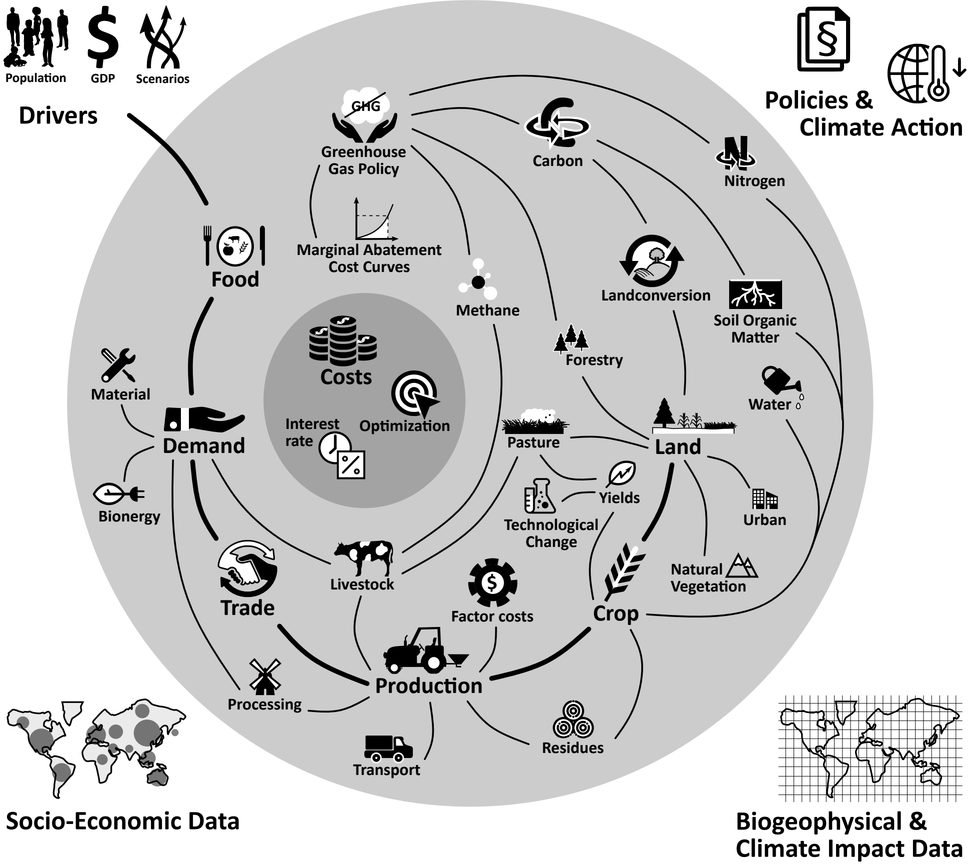
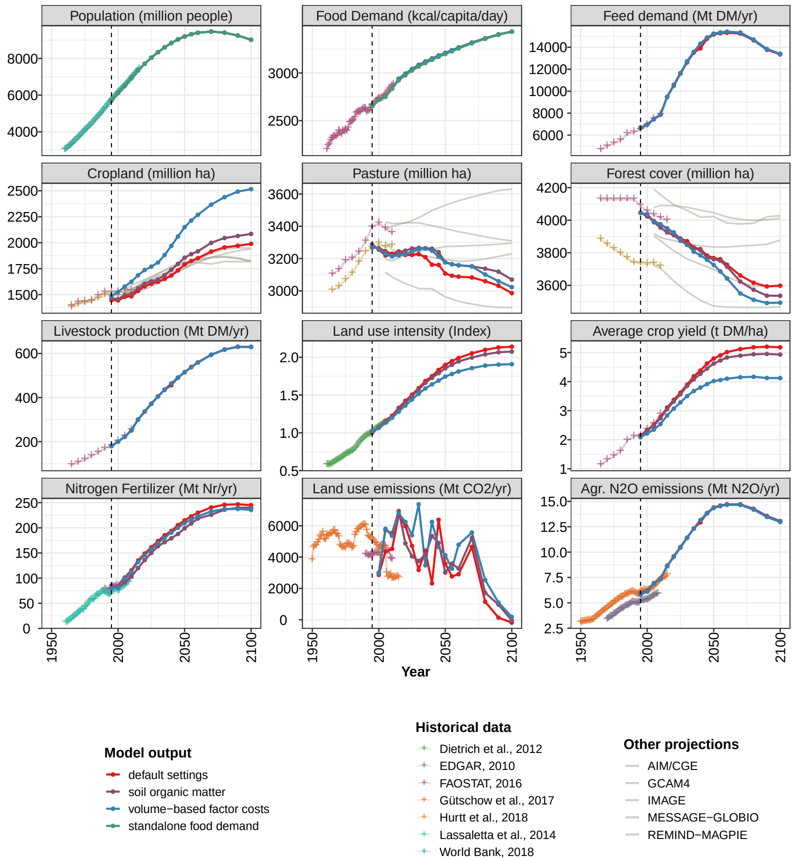

Population
GDP
Scenarios
Drivers
GHG
Greenhouse Gas Policy
Carbon
Nitrogen
Policies & Climate Action
Marginal Abatement Cost Curves
Food
Methane
Landconversion
Soil Organic Matter
Costs
Forestry
Material
Water
Optimization
Interest rate
Demand
Pasture
Land
Yields
Bionergy
Technological Change
Urban
Livestock
Natural Vegetation
Trade
Factor costs
Crop
Production
Processing
Residues
Transport
Socio-Economic Data
Biogeophysical & Climate Impact Data

Fig. 1.



Population (million people)
Food Demand (kcal/capita/day)
Feed demand (Mt DM/yr)
Cropland (million ha)
Pasture (million ha)
Forest cover (million ha)
Livestock production (Mt DM/yr)
Land use intensity (Index)
Average crop yield (t DM/ha)
Nitrogen Fertilizer (Mt Nr/yr)
Land use emissions (Mt CO2/yr)
Agr. N2O emissions (Mt N2O/yr)
Year
Model output
default settings
soil organic matter
volume−based factor costs
standalone food demand
Historical data
Dietrich et al., 2012
EDGAR, 2010
FAOSTAT, 2016
Gütschow et al., 2017
Hurtt et al., 2018
Lassaletta et al., 2014
World Bank, 2018
Other projections
AIM/CGE
GCAM4
IMAGE
MESSAGE−GLOBIO
REMIND−MAGPIE

Fig. 2.