Validation of economic landuse models

1 - Introduction

Land is used for many – often competing – purposes such as food, water, Timber and energy provision, settlements, infrastructure, recreation or biodiversity conservation (e.g.(Lotze-Campen et al., 2010), (Lambin and Meyfroidt, 2011), (Smith et al., 2010),(Smith et al., 2013)). In the future, population and welfare increases and changing consumption patterns are expected to increase the competition for scarce land and water resources and to make it increasingly difficult to fulfill food, water and energy security in the future ((Popp et al., 2011), (Godfray et al., 2010), (Foley et al., 2011)).
To manage these challenges effectively and to find and implement pathways for sustainable land use management that help to deliver vital goods and services, and support human health and well-being, decision makers need to base their decisions on a good understanding of these challenges and the effects of potential reactions to it. To this end, among others, economic landuse models have been developed that assess future pathways of land and water resource usage under different future scenarios. The hope is to get insights that allow for an appropriate societal reaction.
Landuse dynamics in the real world are connected to the ecosystem as well as the economic system, both of which are very complex in nature. Constructing models that produce reliable quantitative results about these dynamics is therefore a very ambitious task. The answer most modelers have to this task is a complex model including a large number of processes as this is perceived to increase the quality of the representation of the real system (Beven, 2002). At the same time, guidelines and standards for testing those models are lacking. It is therefore very difficult for potential users of the model to judge the reliability of the results. Attempts to establish guidelines for good scientific practice in modeling have been made for the field of environmental models (Beven, 2002), (Jakeman et al., 2006), climate models (Knutti, 2008) and recently also in the energy system modeling community ("schwanitz_validation.pdf," n.d.).
One aim of this article is to make a first proposal for a similar set of guidelines for the landuse modeling community. The second section is devoted to this theoretical concept. While agreeing about standards for model validation may in itself be a major step forward, the even bigger challenge is to put this theoretical agreement into effect, i.e. not only to talk about good scientific practice but to change the real processes of model evaluation and development according to those guidelines. Therefore in the third section, we talk about specific tests that can be employed in the evaluation procedure. In the fourth section we present a set of tools that we have developed for validating the economic landuse optimization model MAgPIE (Lotze-Campen et al., 2008), (Popp et al., 2010), (Schmitz et al., 2012). It is designed in a way that it can also be used by other

landuse modeling groups. The fifth section shows some exemplary results of the application of the toolbox to MAgPIE while the sixth and last section contains a conclusion and an outlook to potential future activities.

2 - A theoretical concept for economic landuse model validation

The question how to make sure a theory or model is valid[1] is of major importance in all scientific disciplines and has given rise to a vivid and often controversial debate over the last century. Overviews of the debate from a modeler's perspective can be found in ("schwanitz_validation.pdf," n.d.) and (Barlas and Carpenter, 1990).
Here we want to draw from this pool of ideas to make a proposal for a theoretical foundation of a validation formalism for the economic landuse community.
Firstly, we rely on the widely accepted view of Karl Popper that scientific theories can by no means be verified beyond doubt but should instead be subjected to tests trying to prove the model false (Concept of Falsification) (Popper, 1934). According to Popper, each test that a hypothesis passes, adds to its "Corroboration" thereby increasing the belief in the validity of this hypothesis.

Secondly, we follow the argument by (Barlas, 1996) who claims that system dynamics modeling (of which economic landuse modeling is a sub category) needs a relativist/holistic approach to validation. They argue that "No model can claim absolute objectivity, for every model carries in it the modeler's worldview." And that "Accordingly, model validity is not absolute and validation cannot be entirely objective and formal."

From this, we deduce the following principles that build the foundation of our proposal for a validation concept for the economic landuse model community:

- Validation cannot verify a model it can only increase the belief in its validity. Validation of an economic landuse model should be done by trying "to assess what tests, what trials, it has withstood; that is, we should try to assess how far it has been able to prove its fitness to survive by standing up to tests. (Popper, 1934)
- There is no single objective set of tests that is best suited to increase the belief in the validity of a model. Instead, the set of tests that models are confronted with is and will be based on an (implicit or explicit) agreement within the economic landuse modeling community.

[1] There is a huge debate about the correct name for what we call "valid", "validation" and "validity" (Oreskes et al., 1994). We use those words and hope to be able to define sufficiently what we mean by it.

- “it is impossible to define an absolute notion of model validity divorced from its purpose.” (Barlas, 1996). A model has to be tested for validity for each task it is set to (Sargent, 1998).

From a practical perspective, this means that validation is a continuous effort (“schwanitz_validation.pdf,” n.d.) and that predictions of different states of the system all have to be validated separately. It is therefore crucial to have an automated standard procedure for validation that is easily applicable and allows for validation of a large set of model results in different spatial aggregations.

2.1 - Validation and calibration

We now want to have a look at the close relationship of validation and calibration in the context of economic landuse models.
Calibration of model parameters to make the model usable and accurate is an unavoidable part of model construction, as (Beck et al., 1997) point out.
Whatever validation tests one applies, a validation test that raises concern about the validity of the model, is mostly followed by a calibration procedure that results in a more encouraging test result. We define such a second test result after recalibration as a “posterior validity” test according to (Beck et al., 1997). The authors argue that the fundamental problem with posterior validity is that given the large number of parameters in the model, it is in principle possible to “adjust the behavior of the model so that its match with the behavior of the system may become arbitrarily close. This, however, will provide little insight into how the model will perform under conditions not previously encountered.”(Beck et al., 1997)
In contrast, in order to asses “prior validity”, the test has to be performed and results published without any adjustments to the model based on some previous execution of the same test. These two should not be confused and any attempt to validate a model should try to focus on prior validity[2].

2.2 - Validation is nothing we have to invent – it happens all the time

Judging by published articles (e.g.(Popp et al., 2011),(Lobell et al., 2013),(Sauer et al., 2010)) , economic landuse models are used to deduce from the model results statements about future behavior of the real system. In making these statements, these articles implicitly claim that the models are useful tools to get insights about the real future, i.e. that they are valid for predictions.
A second aspect is that each model needs calibration of its parameters.

[2] Even though (Oreskes et al., 1994) rightly argue that a strict assessment of prior validity is hardly achievable, we argue that the space between prior and posterior validity assessments is a continuum and the aim should be to be “as prior as possible”.

This parameter calibration inevitably involves at least one test of the model behavior (Beck et al., 1997), i.e. a validation test. As the model parameters are calibrated according to the validation test outcome, the resulting validation test shows posterior validity of the model. This posterior validity is then often used as an argument that the model is sufficiently valid to make predictions about the real system which, as discussed above, is not sufficient.
Summing up, economic landuse models are claimed to be valid for predictions. As no clear validation concept in the economic landuse modeling community exists however, validation happens on a rather subjective basis (both, of the modeling teams claiming validity, and the reviewers asserting it).

2.3 - What has to be tested?

Before setting up an evaluation framework consisting of a variety and to a certain extent arbitrary choice of tests, it is important to define first what the tests should aim at. (Barlas, 1996) who is concerned with system dynamics models states that models "must generate the "right output behavior for the right reasons."". We believe that to assure this is also a reasonable goal definition for the set of tests used for economic landuse model validation. So there are two tasks:

1. Testing if a model shows the "right output behavior"
2. Establishing the validity of the models internal structure (Barlas, 1996).

Even though there is a variety of different perspectives on how to validate a model, model output evaluation plays a central role in all the approaches known to us (("schwanitz_validation.pdf," n.d.),(Jakeman et al., 2006),(Knutti, 2008),(Sargent, 1998),(Beven, 2002),(Barlas, 1996)). We will therefore focus our article on task one, testing if a model shows the "right" output behavior. If a model has proven to show the "right" output behavior under different parameterizations and scenarios this may even help to establish the validity of the model's internal structure (Barlas, 1996),(Oreskes et al., 1994). We want to emphasize that other tests and aspects have to be incorporated into the validation formalism ("schwanitz_validation.pdf," n.d.) but with regard to the status quo, facilitating and formalizing the procedure for a thorough output behavior investigation would be a big step forward on the road to establishing good scientific practice in the economic landuse modeling community.

2.4 - Testing if a model shows the "right" output behavior

Testing the output behavior of a model requires prior knowledge or belief about the outputs from independent sources. This can in principle be anything from real world data to expert judgments or other model results. As economic landuse models claim to yield results that give insights about real world processes, we argue that validation against real world data plays a key role in judging if a model produces the "right output

behavior”. There is however a fundamental problem. A characteristic feature of many of these models is that they predict future behavior of the system. (“schwanitz_validation.pdf,” n.d. ,page 9) shows that no experiment can generate future real world data. Therefore direct output validation is rarely possible with economic landuse models. Instead, we propose to perform model output evaluation by comparing model results with historical data. We argue that agreement of a model’s results with real world data in the historical period is a necessary condition to trust the model to perform well in predicting the future. This is what we will call “output validation” throughout this article. Due to the unavoidable extrapolation to the future however, it is not a sufficient condition[3] and other tests have to be performed In addition (Oreskes et al., 1994).
Useful tests include sensitivity analysis and comparison with other model results.

3 - Specific tests

Having established a general approach to validation focusing primarily on output validation, we now want to review some tests that can be applied to test model output behavior.

3.1 - Comparison with historical data

As we argued, comparison with historical data is a key component of the validation process by contributing to task one, testing if the model shows the right output behavior. If prior output validity is tested, it can even be seen as one part of testing the validity of the model’s internal structure (task two). The test should not only aim at comparing the absolute level of the value under consideration between model result and data. The similarity of the trends should also be investigated. Applications can be found in e.g. (Sauer et al., 2010),(Popp et al., 2010),(Jan Philipp Dietrich et al., 2013). There are three major challenges. The first is data quality. Data for validation of economic landuse models is rare and often of questionable quality. Furthermore, mostly no uncertainties of historical data are reported even though they can be assumed to be considerable. Essentially, we have to deal with the following problem (Oreskes et al., 1994):
“What we call data are inference-laden signifiers of natural phenomena to which we have incomplete access”. This is not only true for data about natural phenomena but even more so for socioeconomic data. Therefore comparison with historical data should involve as many independent datasets as possible.

[3] Actually it never is but in this case the shortcoming is even more pronounced.

The second challenge is the usually small overlap between simulation period and historical period. If the overlap is too small, only the difference in absolute level between model results and data can be tested while differences in trend remain hidden. One measure to increase this overlap is to start simulation earlier. Another one is to perform hindcasting analyses as done e.g. in (Lotze-Campen et al., 2008) and (Lobell et al., 2013).
The third challenge is that only one manifestation of the real system exists. Therefore, available historical data is limited and no experiment can be set up to create additional data. As models also need to make use of historical data for calibration, this severely restricts the possibility of assessing prior output validity (see section 'Validation and calibration').

3.2 - Comparison with other model results

Comparison with other model results is often used to underpin the claim of a model's validity (see e.g. (Sauer et al., 2010),("Popp et al. GEC_2010.pdf," n.d.)
But how can comparison with other model results contribute to increasing the belief in the validity of a model?
Let us assume that all the investigated models have been well tested for their output validity for the historical period (A necessary precondition in our opinion). Then, comparing the different future projections will reveal information about the uncertainty in the extrapolation procedure. If the models which all performed well in the past show very different behavior in the future, this is an indication that the extrapolation is uncertain. If however they yield similar results for the future this is unfortunately not a strict proof of "correctness" of the extrapolation because there is still the possibility of a bias common to all models. This possibility decreases however with the number of results of independent models being compared. Therefore, in principle, comparison with other model results can contribute to task one, testing if a model shows the right output behavior. It is however essential to include as many independent projections in the comparison as possible.
Model intercomparison exercises such as AgMIP ("AgMIP," 2013) or ISIMIP (Katja Frieler et al., n.d.) are an example for extensive model output comparisons. It is however often difficult to deduce statements about model validity from such exercises as calibration is involved in the process. For AgMIP, the process can be characterized as follows: All participating models provide a first set of results based on a set of harmonized inputs. These results are made accessible to all participants. Subsequently each group has the chance to remove errors in implementation and recalibrate their model before a second set of results is collected. This procedure can be repeated several times. Therefore, model intercomparison exercises such as AgMIP do not show prior agreement but only posterior (after several iterations of calibration) agreement between model results. The convergence in results that may occur in such model intercomparison exercises therefore does not guarantee an increased validity of model results. It rather reflects a shift from prior to posterior validity results which makes the model outputs less meaningful for model validation among each other. A model intercomparison exercise that aims at

increasing the belief in a model's validity should try to avoid calibration of the participating models based on the comparison.

3.3 - Sensitivity analysis

Another tool that is often applied as a test of economic landuse models is the sensitivity analysis (see e.g. (Schmitz et al., 2012), (Lobell et al., 2013)). It compares model outputs retrieved under different choices for the model's parameters (See (Saltelli et al., 2008) for a detailed description of sensitivity analysis). The range of variation in each parameter is based on known or - more often – assumed uncertainty ranges of this particular parameter. The virtue of sensitivity analyses is to test if there is enough input data of sufficient quality to support the complexity of the model structure (Jakeman et al., 2006). If this is not the case, the parameters are not well constrained, sensitivity will be high and the results contain little information. In such a case, a simpler model with a better constrained parameter space may be preferable (Marsili-Libelli and Checchi, 2005). The major practical challenge with sensitivity analyses is the often large number of model parameters which makes it impossible or at least very cumbersome to explore the whole space of model parameterizations during the test.

4 – Our approach to output validation

In the previous section we argued that model output validation is one of the pillars of testing model validity. In this section, we present a set of tools that we developed for output validation. Only if a standardized and user friendly set of tests is available, output validation can be carried out for a variety of outputs and for each new development stage of the model. Our toolbox has been developed for validation of the economic landuse model MAgPIE, a global land use allocation model. It takes regional economic conditions as well as spatially explicit data on potential crop yields, land and water constraints into account and derives specific land use patterns, water withdrawals and emissions by minimizing global production costs.
We have designed our toolbox in a way that makes it also applicable for output validation of other landuse models. The toolbox is based on R, an open source language for statistical computing and graphics (Team, 2012).

4.1 - Database

As discussed, model output validation comprises comparison with historical data as well as comparison with other model results.

The core of our toolbox is a database for collection of historical data as well as independent model projections. The aim is to collect as many datasets as possible for

each output under consideration in order to include the largest possible amount of information in the output validation. Each dataset entry is a time series with central value and upper and lower uncertainty bounds if available.
The entries in the database are grouped according to output name (e.g. “cropland”), data type (historical data or projection) and a name for the dataset based on the source (e.g. “FAO”). Additionally, information about the unit and the source of a dataset is stored. Supported aggregation levels are 0.5 degree, country (defined by the ISO 3166-1 alpha3 code (“ISO 3166-1 alpha-3,” n.d.)) and global. Data on Continent or region level is excluded because these definitions are ambiguous. For each dataset, a method for spatial aggregation (e.g. sum, mean) has to be specified.
The database includes a function for user friendly data retrieval. The user only has to specify the desired output (e.g. cropland) and the desired spatial aggregation. The function will check the database for all datasets for this output that are spatially at least as disaggregated as the user defined aggregation level. It then returns all the matching datasets aggregated to the desired level.
A key component of the database is a user friendly function for incorporation of new datasets. This facilitates inclusion of new or newly discovered datasets by all users and makes sure that the database gets more complete and useful over time. Several consistency checks of newly entered data are performed to assure that the dataset is provided in the correct format and all necessary information is provided.

4.2 - Validation Plot

Comparison of model outputs with historical data and other projections can be done by several means. A widely applied one is visual comparison in a plot. Our toolbox includes a function to easily produce such validation plots. The model output under consideration as well as historical data and other projections from the database are time series. Therefore the validation Plot depicts the quantity under consideration over time. In the graph, there is a clear optical distinction between the model output to be validated, historical trends and other projections. It also includes a legend with the names of all datasets. For country level and cellular data, this style of plot does not make sense.

5 - Results

In this section we want to show a few exemplary validation plots obtained with our toolbox. Note that all the plots have directly been derived from the toolbox and no further processing has taken place, i.e. we can reproduce these plots easily for new model development stages and across a large number of simulations.
Figure 1 shows an example of the plotting routine of our toolbox. MAgPIE global CO_2 emissions from landuse change are compared to FAO (FAOSTAT. 2013, n.d.) data. The heading and the y label are automatically retrieved from the database. There is a clear visual distinction between the different components of the plot. The model output that

is subjected to the test is always depicted in red with solid dots while fainter solid lines with empty circles are used for historical data. A vertical dashed line represents the start of the simulation period. The legend has subtitles indicating whether a dataset is model output to be validated, other projection or historical data.

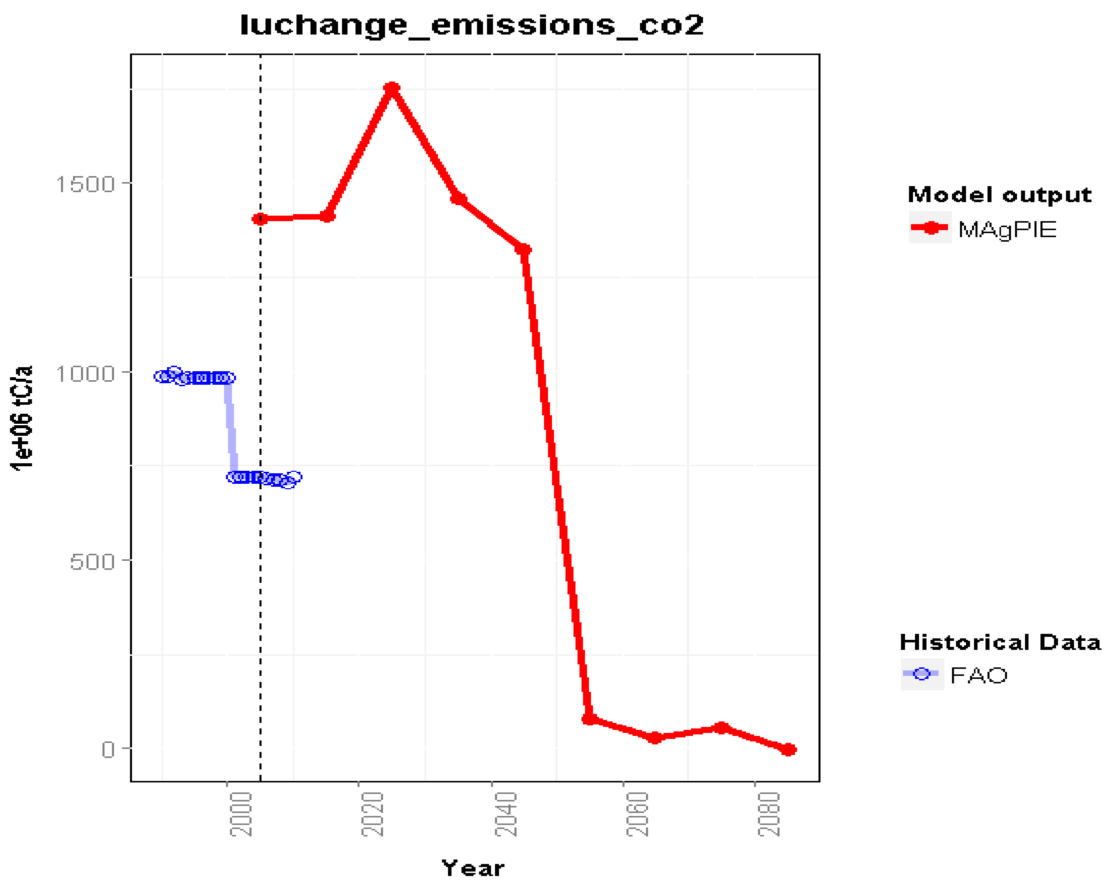

Figure 1: MAgPIE global CO2 emissions from landuse change compared to FAO (FAOSTAT. 2013, n.d.) data.

Figure 2 shows the same MAgPIE result this time compared to all the historical datasets currently available in the database. Uncertainty ranges are depicted as grey shaded error regions.

Comparing the two figures shows that including more independent information on global landuse change emissions increases the perceived uncertainty. This in return highlights that by not including this additional information a false perception of data accuracy is conveyed which will then result in models falsely failing the visual output validation test and finally in over calibration of models (Jakeman et al., 2006).

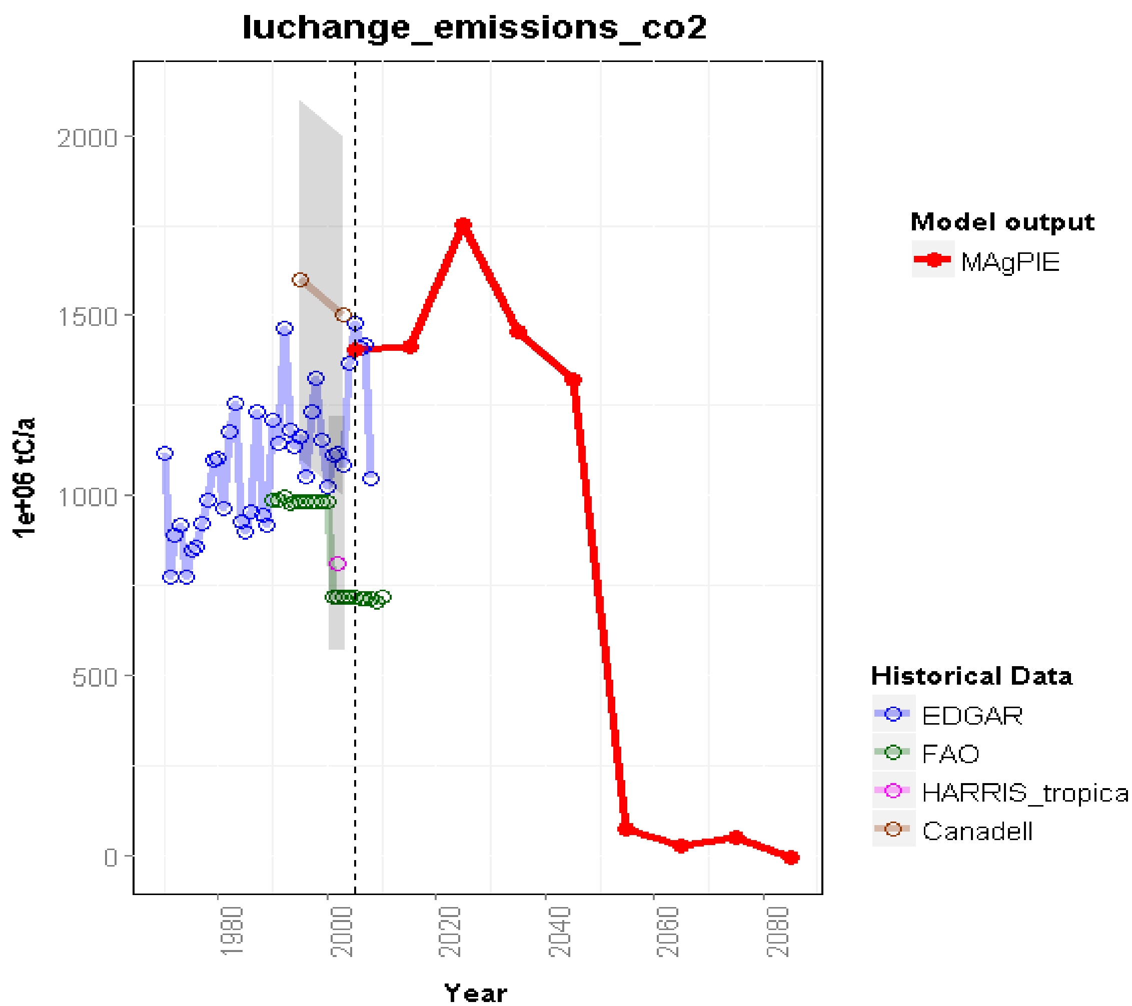

Figure 2 : MAgPIE Global CO2 emissions from landuse change compared to EDGAR (European Commission, 2010), FAO (FAOSTAT. 2013, n.d.) , HARRIS_tropical (Harris et al., 2012) and Canadell (Canadell et al., 2007) data.

Figure 3 is an example of a spatially more disaggregated validation plot. In this case, global CO2 emissions from MAgPIE due to land use change are compared to data on the level of the 10 MAgPIE world regions. Comparing this to Figure 2 shows that the estimates from the sources HARRIS_tropical (Harris et al., 2012) and Canadell (Canadell et al., 2007) drop out because they only estimate emissions on a global level.

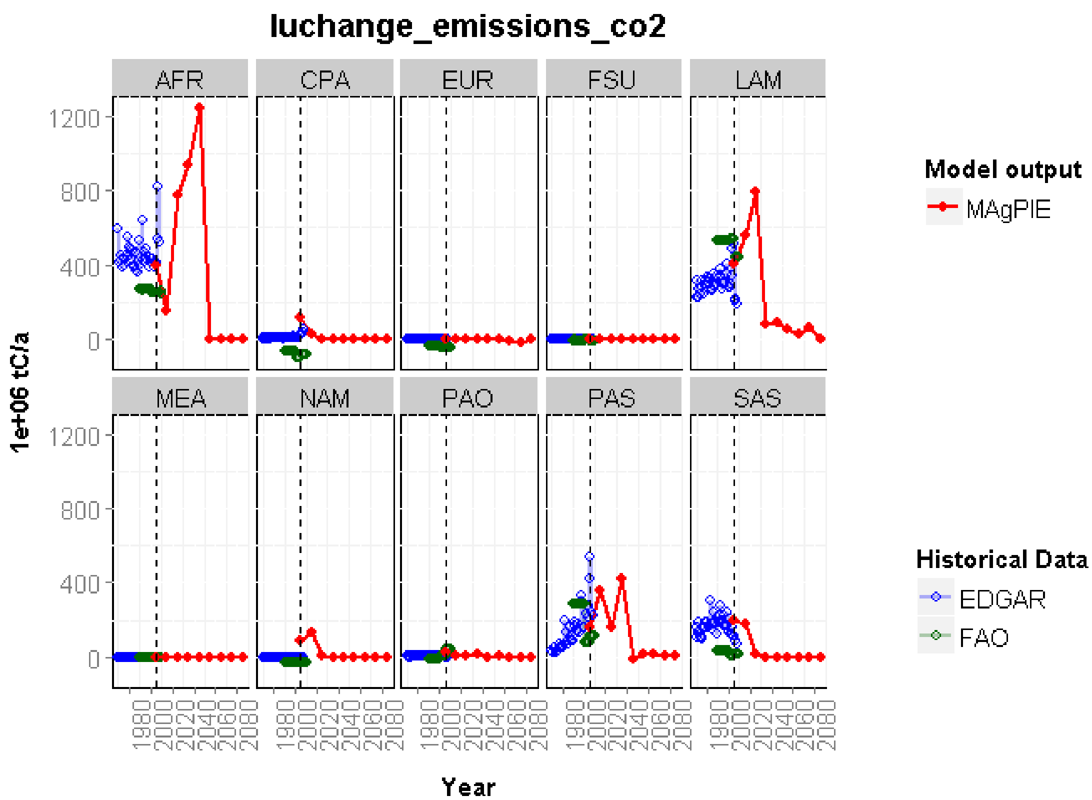

Figure 3: MAgPIE regional CO2 emissions from landuse change compared to EDGAR (European Commission, 2010) and FAO (FAOSTAT. 2013, n.d.) , data.

Figure 3 shows an example of a validation plot involving comparison with historical data as well as with independent model projections. The model output under investigation is area actually irrigated. Other projections are displayed as faint dashed lines with faint solid dots in order to make them clearly distinguishable from historical data and model output under investigation. Uncertainty ranges can also be included for projections. In this case the uncertainty range of the WATERGAP projections (ALCAMO_WATERGAP dataset, (Alcamo et al., 2000)) is based on projections under different scenario assumptions and is displayed as a grey shaded region just as uncertainties of historical data. The spread among the four different projections is considerable although the starting points of all projections are quite similar. This may be a sign of common calibration to a specific historical dataset. Uncertainty in data, for the year 2000 according to (Siebert and Döll, 2002) is higher than the spread in models at that point of time. The MAgPIE projection starts significantly lower than the other estimates and is at the lower end of the range of historical data in the year 2000. This is due to the fact that MAgPIE only considers renewable water resources.

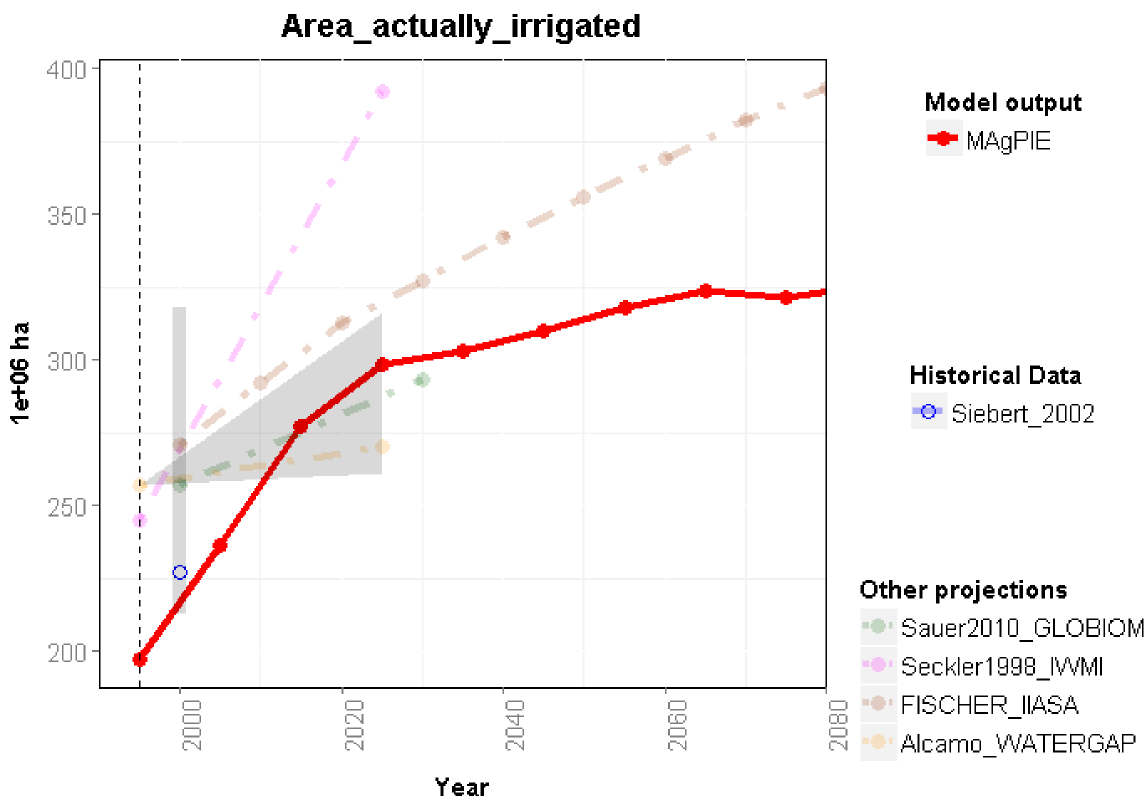

Figure 4: MAgPIE global irrigated area compared to data from (Siebert and Döll, 2002) and projections from (Sauer et al., 2010), (Seckler, 1998), (Fischer and Tubiello, 2007), (Alcamo et al., 2000)

Table 1 finally shows the current content of our database.

	Current content of the database		
Type of data (name in the database)	Unit	Number of historical datasets	Number of projections
cropland_physical	10^6 ha	2	0
Pastureland	10^6 ha	2	0
luchange_emissions_co2	10^6 tC/a	4	0
Area_equipped_for_irrigation	10^6 ha	1	0
Area_actually_irrigated	10^6 ha	1	4
Agricultural_water_withdrawals	km^3/a	1	4

Table 1: Current content of our validation database. Type of data indicated by name in the database, unit, Number of historical datasets included and number of other projections included for each dataset.

6 - Conclusion

In the second section we have made a proposal for a theoretical framework of landuse model validation. The key points are:

- Validation is a continuous effort; validation exercises have to be performed according to the application of the model and for each new development stage
- There is no objective notion of model validity. A community agreement on validity criteria has to be found.
- Validation should be strictly separated from calibration; validation exercises should aim at testing prior validity, i.e. should be based on independent information that has not been used in the calibration process.
- Validity tests should include tests of the output behavior as well as the internal structure of a model. To begin with, it is more important to focus on testing the output behavior because this is already being done and can easily be improved.

In the third section we have reviewed three tests that are already frequently executed and proposed improvements:

- **Comparison with historical data** is the most important test of model validity. Due to the lack of accurate historical data, as many independent datasets as available should be included in the comparison. In order to be able to compare trends and not only single values, the overlap between historical data and model results has to be as large as possible. This can be achieved by starting simulation earlier.

- **Comparison with other model results** are an important test of a model's output behavior. In order to avoid bias, it is necessary to include as many independent projections as possible in the comparison. Calibration during the comparison should be avoided in order to make a statement about prior validity which is by far stronger than statements about posterior validity that can be obtained if calibration is involved in the process.
- **Sensitivity analyses** are an important test of the validity of the model structure. For most models, it is however hardly possible to perform a full sensitivity analysis due to the large number of model parameters

In the fourth section we have presented a toolbox for validation. It facilitates the comparison of model results with historical data as well as comparison with other model results. The core is a flexible and growing database together with a function to create plots for visual output validation of model results against all information in the database. In the fifth section we have shown an application of the toolbox to MAPIE results that highlights the importance of collecting as much independent information as possible for comparison.

Further development of the toolkit will include the incorporation of new datasets as well as the development of an algorithm for automatic output validation based on statistical tests of the similarity between model output trends and trends in data. We would like to make the toolkit openly accessible to the whole community and hope that others will join in the development in order to make it more useful and complete.

References

AgMIP [WWW Document], 2013. . URL http://www.agmip.org/about-us/

Alcamo, J., Henrichs, T., Rösch, T., 2000. World Water in 2025 – Global modeling and scenario analysis for the World Commission on Water for the 21 st Century. Report A0002.

Barlas, Y., 1996. Formal aspects of model validity and validation in system dynamics. System dynamics review 12, 183–210.

Barlas, Y., Carpenter, S., 1990. Philosophical roots of model validation: Two paradigms. System Dynamics Review 6, 148–166.

Beck, M.B., Ravetz, J.R., Mulkey, L.A., Barnwell, T.O., 1997. On the problem of model validation for predictive exposure assessments. Stochastic Hydrology and Hydraulics 11, 229–254.

Beven, K., 2002. Towards a coherent philosophy for modelling the environment. Proceedings of the Royal Society of London. Series A: Mathematical, Physical and Engineering Sciences 458, 2465–2484.

European Commission, 2010. Emission Database for Global Atmospheric Research (EDGAR), release version 4.2.

FAOSTAT. 2013, n.d. Database Collection of the Food and Agriculture Organization of the United Nations.

Fischer, K., Tubiello, B., 2007. Climate change effects on irrigation water requirements: Effects of mitigation, 1990 - 2080.

Foley, J.A., Ramankutty, N., Brauman, K.A., Cassidy, E.S., Gerber, J.S., Johnston, M., Mueller, N.D., O'Connell, C., Ray, D.K., West, P.C., Balzer, C., Bennett, E.M., Carpenter, S.R., Hill, J., Monfreda, C., Polasky, S., Rockström, J., Sheehan, J., Siebert, S., Tilman, D., Zaks, D.P.M., 2011. Solutions for a cultivated planet. Nature 478, 337–342.

Godfray, H.C.J., Beddington, J.R., Crute, I.R., Haddad, L., Lawrence, D., Muir, J.F., Pretty, J., Robinson, S., Thomas, S.M., Toulmin, C., 2010. Food Security: The Challenge of Feeding 9 Billion People. Science 327, 812–818.

ISO 3166-1 alpha-3, n.d. .

Jakeman, A.J., Letcher, R.A., Norton, J.P., 2006. Ten iterative steps in development and evaluation of environmental models. Environmental Modelling & Software 21, 602–614.

Jan Philipp Dietrich, Christoph Schmitz, Hermann Lotze-Campen, Alexander Popp, Christoph Mueller, 2013. Forecasting technological change in agriculture - An endogenous implementation in a global land use model. Technological Forecasting and Social Change.

Knutti, R., 2008. Should we believe model predictions of future climate change? Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences 366, 4647–4664.

Lambin, E.F., Meyfroidt, P., 2011. Inaugural Article: Global land use change, economic globalization, and the looming land scarcity. Proceedings of the National Academy of Sciences 108, 3465–3472.

Lobell, D.B., Baldos, U.L.C., Hertel, T.W., 2013. Climate adaptation as mitigation: the case of agricultural investments. Environmental Research Letters 8, 015012.

Lotze-Campen, H., Müller, C., Bondeau, A., Rost, S., Popp, A., Lucht, W., 2008. Global food demand, productivity growth, and the scarcity of land and water resources: a spatially explicit mathematical programming approach. Agricultural Economics 39, 325–338.

Lotze-Campen, H., Popp, A., Beringer, T., Müller, C., Bondeau, A., Rost, S., Lucht, W., 2010. Scenarios of global bioenergy production: The trade-offs between agricultural expansion, intensification and trade. Ecological Modelling 221, 2188–2196.
Marsili-Libelli, S., Checchi, N., 2005. Identification of dynamic models for horizontal subsurface constructed wetlands. Ecological Modelling 187, 201–218.
Oreskes, N., Shrader-Frechette, K., Belitz, K., 1994. Verification, Validation, and Confirmation of Numerical Models in the Earth Sciences. Science 263, 641–646.
Popp, A., Dietrich, J.P., Lotze-Campen, H., Klein, D., Bauer, N., Krause, M., Beringer, T., Gerten, D., Edenhofer, O., 2011. The economic potential of bioenergy for climate change mitigation with special attention given to implications for the land system. Environmental Research Letters 6, 034017.
Popp, A., Lotze-Campen, H., Bodirsky, B., 2010. Food consumption, diet shifts and associated non-CO2 greenhouse gases from agricultural production. Global Environmental Change 20, 451 – 462.
Popp et al. GEC_2010.pdf, n.d. .
Popper, K.R., 1934. The Logic of Scientific Discovery. Hutchinson, London.
Saltelli, A., Ratto, M., Andres, T., Campolongo, F., Cariboni, J., Gatelli, D., Saisana, M., Tarantola, S., 2008. Global Sensitivity Analysis. The Primer. John Wiley & Sons, Ltd.
Sargent, R.G., 1998. Verification and validation of simulation models, in: Proceedings of the 30th Conference on Winter Simulation. pp. 121–130.
Sauer, T., Havlík, P., Schneider, U.A., Schmid, E., Kindermann, G., Obersteiner, M., 2010. Agriculture and resource availability in a changing world: The role of irrigation. Water Resources Research 46, W06503.
Schmitz, C., Biewald, A., Lotze-Campen, H., Popp, A., Dietrich, J.P., Bodirsky, B., Krause, M., Weindl, I., 2012. Trading more food: Implications for land use, greenhouse gas emissions, and the food system. GLOBAL ENVIRONMENTAL CHANGE-HUMAN AND POLICY DIMENSIONS 22, 189–209.
schwanitz_validation.pdf, n.d. .
Seckler, D.W., 1998. World Water Demand and Supply, 1990 to 2025: Scenarios and Issues. IWMI.
Siebert, S., Döll, P., 2002. Global modeling of irrigation water requirements. Water Resources Research 38.
Smith, P., Gregory, P.J., van Vuuren, D., Obersteiner, M., Havlik, P., Rounsevell, M., Woods, J., Stehfest, E., Bellarby, J., 2010. Competition for land. Philosophical Transactions of the Royal Society B: Biological Sciences 365, 2941–2957.
Smith, P., Haberl, H., Popp, A., Erb, K., Lauk, C., Harper, R., Tubiello, F., de Siqueira Pinto, A., Jafari, M., Sohi, S., Masera, O., Böttcher, H., Berndes, G., Bustamante, M., Ahammad, H., Clark, H., Dong, H., Elsiddig, E.A., Mbow, C., Ravindranath, N.H., Rice, C.W., Robledo-Abad, C., Romanovskaya, A., Sperling, F., Herrero, M., House, J.I., Rose, S., 2013. How much land based greenhouse gas mitigation can be

achieved without compromising food security and environmental goals? Global Change Biology n/a–n/a.
Team, R.C., 2012. R: A Language and Environment for Statistical Computing. Vienna, Austria.