A framework to assess model structure stationarity

J. A. Verstegen[1], F. van der Hilst[1], D. Karssenberg[2]

[1]Copernicus Institute for Sustainable Development and Innovation, Faculty of Geosciences, Utrecht University, Heidelberglaan 2, 3584 CD Utrecht, The Netherlands,
Telephone: (+31)302533544
Email: J.A.Verstegen@uu.nl

[2]Department of Physical Geography, Faculty of Geosciences, Utrecht University, Heidelberglaan 2, PO Box 80115, 3508 TC Utrecht, The Netherlands

Abstract

In geosimulation models it is often, implicitly or explicitly, assumed that the relationship between the system state change and its explanatory processes, i.e. model structure, is stationary. Potential systemic changes influencing the stationarity of this relationship are thereby disregarded. We present a framework to assess model structure stationarity and apply it on a case study in which we model land use change in São Paulo state, Brazil, from 2004 to 2012. We find that systemic changes do occur during this period, which is a warning that the assumption of stationarity is likely to cause an underestimation of simulated future uncertainty.

1. Introduction

Building a geosimulation or spatio-temporal model involves 1) creating a model structure, i.e. identifying the relevant processes in the system plus the way they interact and implementing these as transition rules in the model, and 2) parameterizing the model, i.e. setting the model parameters so that the processes are correctly represented for a given case study. Ideally, a geosimulation model is fully stochastic, so that uncertainty in the simulated system state change can be quantified (Heuvelink, 1998). Building a fully stochastic model requires having a preselection of potential model structures instead of a single best-guess model structure, as well as a prior probability distribution for each parameter instead of a single value. By assimilating a dataset of observations of the modelled system the likelihood of each model structure being correct can be assessed and the posterior probability distribution for each parameter can be determined (e.g., Verstegen et *al.*, 2014).

Once such a stochastic model has been set up, modelers often assume that the established relationship between the system state change and its explanatory processes is stationary. Accordingly, model structure and parameterization are usually kept constant over model runtime, disregarding potential systemic changes in this relationship resulting from e.g., climate change or societal transformations. If systemic changes are found to have occurred in the past, simulated future uncertainty is probably underestimated (Verstegen et *al.*, in press).

Here, we present a framework that can 1) estimate posterior probabilities of model structures and parameters, given a range of potential model structures, prior probability distributions of all stochastic parameters and a dataset of observed system state changes over time, and 2) assess the stationarity of this system. We test the framework by modeling land use change in São Paulo state, Brazil, using a Cellular Automata (CA) type of model.

2. General framework

For a general geosimulation model, potential model structures and prior probability distributions for each parameter can be estimated by expert judgment or literature review. Computationally, the stochastic structure and parameters can be represented by running a Monte Carlo simulation, where each ensemble member *i*, with *i* = 1, 2, …, *I*, is created by obtaining a sample from the potential structures and each of these probability distributions. In this way a stochastic model is obtained in which the defined errors propagate to the modelled system state.

We use a Bayesian estimation technique, the particle filter (van Leeuwen, 2009), to reduce the uncertainty in the ensemble based on observations of the real system state. At a time step when observations are available, the particle filter is applied:

$$p\left(\mathbf{z}_t^i \middle| \mathbf{o}_t\right) = \frac{p\left(\mathbf{o}_t \middle| \mathbf{z}_t^i\right) \cdot p\left(\mathbf{z}_t^i\right)}{p(\mathbf{o}_t)} = \frac{p\left(\mathbf{o}_t \middle| \mathbf{z}_t^i\right) \cdot p\left(\mathbf{z}_t^i\right)}{\sum_{j=1}^{N} p\left(\mathbf{o}_t \middle| \mathbf{z}_t^j\right) \cdot p\left(\mathbf{z}_t^j\right)}, \text{ for each } i = 1, 2, \ldots, I \qquad (1)$$

In equation 1, $p\left(\mathbf{z}_t^i\right)$ is the prior probability of the model state of ensemble member *i*. The model state consists of the system state(s) as well as the transition rules, inputs and parameters. Furthermore, $p(\mathbf{o}_t)$ is the probability distribution of the observations, i.e. the measurement data and their uncertainty, and $p\left(\mathbf{z}_t^i \middle| \mathbf{o}_t\right)$ is the posterior probability of the model state $\mathbf{z}_t^i$ of *i*, given these observations. Finally, $p\left(\mathbf{o}_t \middle| \mathbf{z}_t^i\right)$ is the probability of the observations given member *i*; $p\left(\mathbf{o}_t \middle| \mathbf{z}_t^i\right)$ can be determined by a function of the difference between the modelled system state and the observed system state of *i*. This is further explained in Verstegen et *al.* (2014, in press).

Equation 1 can be evaluated in two ways: at each time step *t* when observations are available $p\left(\mathbf{z}_t^i\right)$ can be either 1) the chosen prior probability distribution, or 2) the posterior probability distribution from the previous moment that observations were available, i.e. the particle filter is applied sequentially. In method 1, distinctive information about $p\left(\mathbf{z}_t^i \middle| \mathbf{o}_t\right)$ at that point in time is obtained, while in method 2, additive information is obtained. Method 2 is the traditional approach of the particle filter, in which the probability distributions of model structure and parameters gradually converge to their best value. Yet, this approach should only be applied when it can be assumed that model structure and parameters are stationary, because if they are not, there is no single model structure or single parameter the model state can converge towards.

Therefore, we apply the particle filter in both ways, method 2, hereafter called 'traditional approach', to estimate the best-guess posterior probabilities of model structures and parameters over all observation moments, and method 1, hereafter called 'systemic change test', to assess to what extent the posterior probability distributions found at the different observation moments vary over time. Using statistical tests, we check if potential variations in the distributions are significant and if their means show a temporal pattern (see Verstegen et *al.*, in press).

3. Case study

A land use change Cellular Automaton is set up to test our framework. The model simulates the expansion of sugar cane in São Paulo state, Brazil, on a 5 km resolution raster. Four processes that might be of relevance in where sugar cane expands are implemented: number of sugar cane cells in an extended Moore neighborhood, distance to sugar cane mills, potential yield of the land, and slope. Weights between zero and one are assigned to these processes, determining their relative importance in the total transition rule. These weights have an a priori uninformed distribution, and are a posteriori determined by the particle filter, using observations from Rudorff et *al.* (2010) from 2004 to 2012.

4. Results

Using the traditional particle filter method we find that the mean weights of the processes determining where sugar cane expands are 0.23 for sugar cane in the neighborhood, 0.36 for distance to mills, 0.13 for potential yield, and 0.28 for slope (Figure 1). The posterior distribution in the traditional approach of the weight for sugar cane in the neighborhood clearly shows two peaks, suggesting that there is no single model structure the model state can converge towards.

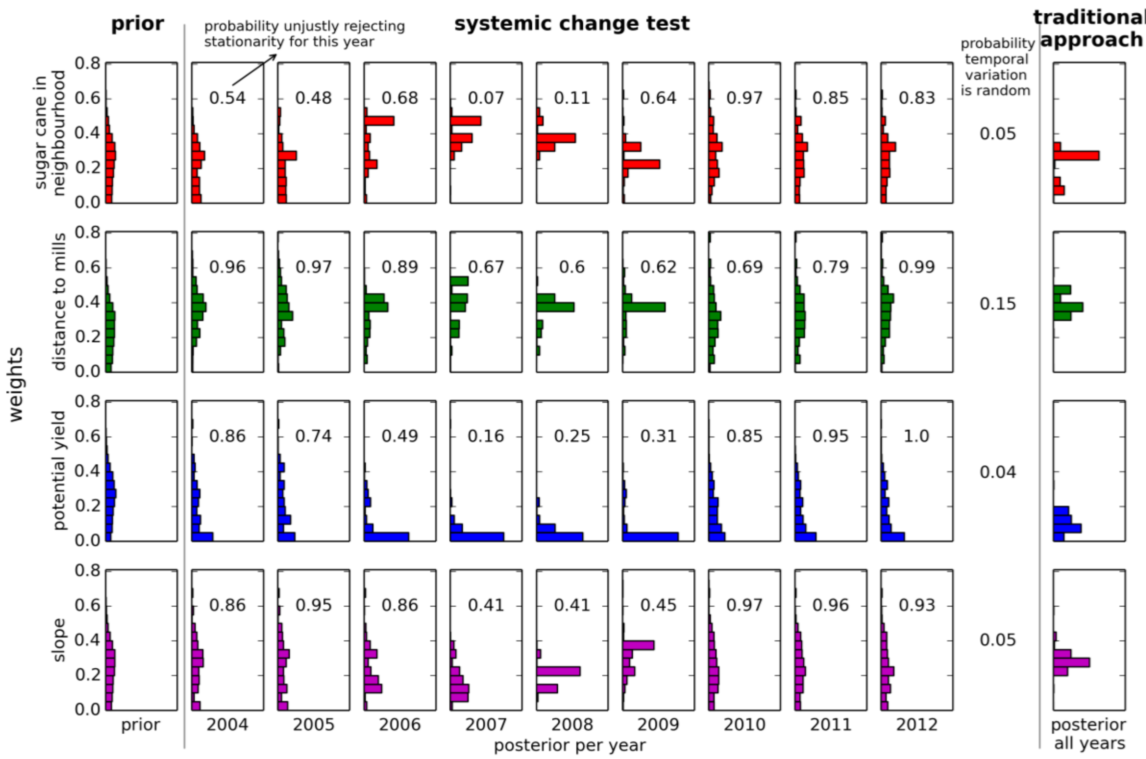

Figure 1: Prior (left) and posterior distributions of the weights of sugar cane in the neighborhood, distance to mills, potential yield, and slope, obtained with the 'traditional approach' (right) and with the 'systemic change test' (middle). The values inside the bar plots indicate the probability that the null hypothesis of stationarity is unjustly rejected for that year. The value on the right gives the probability that the pattern found in the deviations from the mean is random. Hence, for both tests: low values indicate a high

probability of systemic change and high values indicate a high probability of stationarity. Adapted from Verstegen et al. (in press).

The systemic change test confirms this suggestion (Figure 1): the posterior distributions of sugar cane in the neighborhood in especially 2007 and 2008 are significantly different from the rest. The probability that the null hypothesis of stationarity is unjustly rejected for these years is low, 0.07 and 0.11, indicating that the change is significant (cf., Verstegen et *al.*, in press). Also, the probability that the temporal variation in the mean is random is low, 0.05, signifying systemic change with a temporal pattern. For the weights of potential yield and slope the same reasoning is true, but with lower significance levels. The weight of distance to mills seems to be stationary, but still the overall model structure (combination of all process) is not.

5. Discussion and conclusions

We have presented a framework that can 1) estimate posterior probabilities of model structures and parameters of a geosimulation model based on observed system state changes over time, and 2) assess the stationarity of this system. The advantages and drawbacks of the framework are discussed in Verstegen et *al.* (2014, in press). In our case study the assessed model structure was found to be non-stationary; a warning for the modeler to be careful drawing conclusions from simulated projections or even not to use the model for future projections.

6. Acknowledgements

This work was carried out within the BE-Basic R&D Program, which was granted a FES subsidy from the Dutch Ministry of Economic affairs, agriculture and innovation (EL&I).

7. References

Heuvelink, G.B.M., 1998. *Error Propagation in Environmental Modelling with GIS*. Taylor & Francis, London, United Kingdom.

Rudorff, B.F.T., Aguiar, D.A., Silva, W.F., Sugawara, L.M., Adami, M., Moreira, M.A., 2010. Studies on the Rapid Expansion of Sugarcane for Ethanol Production in São Paulo State (Brazil) Using Landsat Data. Remote Sensing 2: 1057-1076.

van Leeuwen, P.J., 2009. Particle filtering in geophysical systems. Monthly Weather Review 137: 4089-4114.

Verstegen, J.A., Karssenberg, D., van der Hilst, F., Faaij, A.P.C., in press. Detecting systemic change in a land use system by Bayesian data assimilation. *Environmental Modelling & Software*.

Verstegen, J.A., Karssenberg, D., van der Hilst, F., Faaij, A.P.C., 2014. Identifying a land use change cellular automaton by Bayesian data assimilation. *Environmental Modelling & Software* 53: 121-136.