Environmental Modelling & Software 53 (2014) 121–136

Contents lists available at ScienceDirect

Environmental Modelling & Software

journal homepage: www.elsevier.com/locate/envsoft

Identifying a land use change cellular automaton by Bayesian data assimilation

Judith A. Verstegen [a,*], Derek Karssenberg [b], Floor van der Hilst [a], André P.C. Faaij [a]

[a] Copernicus Institute for Sustainable Development and Innovation, Faculty of Geosciences, Utrecht University, Heidelberglaan 2, 3584 CS Utrecht, The Netherlands

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

ARTICLE INFO

Article history:
Received 25 January 2013
Received in revised form
15 November 2013
Accepted 28 November 2013
Available online 20 December 2013

Keywords:
Data assimilation
Cellular automata
Calibration
Model structure
Land use change
Particle filter

ABSTRACT

We present a Bayesian method that simultaneously identifies the model structure and calibrates the parameters of a cellular automaton (CA). The method entails sequential assimilation of observations, using a particle filter. It employs prior knowledge of experts to define which processes might be important in the system, and uses empirical information from observations to identify which ones really are and how these processes should be parameterized. In a case study for the São Paulo state in Brazil, we identify a land use change CA simulating sugarcane cropland expansion from 2003 to 2016. Eight annual observation maps of sugar cane cultivation are used, split over space and time for calibration and validation. It is shown that the identified CA can properly reproduce the observations, and has a minimum reduction factor of 3 in root mean square error compared to a Monte Carlo simulation without particle filter. In the part of the study area where no observational data are assimilated (validation area), there is little reduction in model performance compared to the part with observational data. So, incomplete datasets, regional land survey data, or clouded remote sensing images can still provide useful information for this particle filter method, which is an advantage because good quality land use maps are rare. Another advantage is that in our approach the output uncertainty encompasses errors from expert knowledge, model structure, parameters and observation (calibration) data. This can, in our opinion, be very useful for example to determine up to what future period the results are a secure basis for decisions and policy making.

1. Introduction

A Cellular Automaton (CA) represents spatio-temporal change as local interactions of different entities and processes in a raster environment (Santé et al., 2010). The fact that a CA consists of relatively simple rules that can lead to complex patterns, makes it suitable to study complex system behaviour, which is currently considered important in environmental systems research (Page, 2011; Manson, 2007; Johnson, 2010; Grimm and Railsback, 2012). Therefore, cellular automata are applied in many environmental modelling domains, like fire propagation (Berjak and Hearne, 2002), vegetation spreading (Kéfi et al., 2007), and urban or land use change modelling (Verburg et al., 2004; Batty, 2005; Lauf et al., 2012). In CA development, one can distinguish between model structure identification, i.e. finding the set of processes to be represented in the model, conceptualized into the set of transition rules, and model calibration, i.e. finding the correct parameterization of these processes. In urban and land use change modelling, finding the set of transition rules is problematic (Santé et al., 2010; Straatman et al., 2004), which possibly poses limitations on the reliability and therefore the usability of these models.

Transition rule derivation can be done in a number of ways. 1) From fundamental, e.g., physical or chemical, laws (e.g., Collin et al., 2011). This is difficult in land use change modelling, as most fundamental laws in this field do not provide a quantitative process description. Yet, some have successfully applied physical laws to simulate land use expansion, mainly aimed at cities (Batty, 2012; Bettencourt, 2013). 2) By experts, who have experience-based knowledge of the study area. This is widely done in land use change modelling (e.g., van der Hilst et al., 2012; Yu et al., 2011), but it is somewhat subjective. 3) From empirical data. It is recognized that this is challenging in land use change modelling (Straatman et al., 2004; Hansen, 2012), but it is still important to continue exploring this option, because there is a need to find a more evidence-based approach to set up a land use change model.

* Corresponding author.
E-mail address: J.A.Verstegen@uu.nl (J.A. Verstegen).

1364-8152/$ – see front matter © 2013 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.envsoft.2013.11.009

One can combine the benefits of expert knowledge and empirical data by using a method for transition rule derivation in which the prior knowledge (definition of potential model structures) is defined by experts, and the posterior knowledge (identification of the best structure) is attained by empirical data. Our objective is to devise such a method, which we believe should fulfil two requirements. The first requirement is that the method should be able to quantify uncertainty (Rasmussen and Hamilton, 2012; Aerts et al., 2003), i.e. it should not only be able to select the best model structure from all potential model structures defined by the prior knowledge, but it should give the likelihood of each individual structure being correct. In this way, a stochastic CA is obtained, which combines all potential model structures and parameters in an optimal way. The most important advantage of this is that confidence intervals of the modelled land use projections can be defined, such that policy makers can decide up to what point in time the projections are reliable enough to be a foundation for their policies. The second requirement is that, herein, one should not only take into account uncertainty in the prior information, but also in the empirical data, the observations of land use, used to update the priors (Fang et al., 2006). Ignoring uncertainty in the empirical data may lead to an underestimation of model output uncertainty.

The combined requirements of prior knowledge, observation uncertainty, and posterior knowledge with output uncertainty lead towards Bayesian methods, which start out with prior knowledge, and then assemble model uncertainty and observation uncertainty to end up with posterior knowledge including uncertainty information. Therefore, we show a method for model structure identification and calibration using the particle filter, a sequential Bayesian estimation, or data assimilation, technique (van Leeuwen, 2009). Data assimilation techniques update the prior knowledge during model runtime at time steps when observations are available. We will use this property to sequentially update both the model rules and their parameters. Data assimilation techniques are increasingly being used to calibrate spatio-temporal models in a wide range of different fields in the environmental sciences, such as oceanography (van Leeuwen, 2003), hydrology (Salamon and Feyen, 2009), and atmospheric transport (Hiemstra et al., 2012), but have, to our knowledge, not yet been applied for model structure identification. Recently, their potential has been recognized in the land use change field (van der Kwast et al., 2011; Zhang et al., 2011).

The approach that is most often used in land use change modelling to define the model structure is regression on a land use map (Verburg et al., 1999, 2002; Aguiar et al., 2007; Diogo et al., in preparation). This method mostly results in only one deterministic model structure, without uncertainty in either the observations used to construct the regression model or in the model itself, and therefore does not meet our requirements. In the last decade, model rule identification methods originating from artificial intelligence have become popular, like neural networks (Dai et al., 2005; Li and Yeh, 2002), and swarm intelligence algorithms (Feng et al., 2011; Liu et al., 2008). These, however, do not take into account observation uncertainty, the second requirement. Moreover, they result in black-box models (Li and Yeh, 2002), i.e. they do not provide explicit posterior knowledge. Bayesian land use model structure identification has been performed before by Kocabas and Dragićević (2007). They apply a Bayesian network and an influence diagram. However, they do not include observation uncertainty.

In this study, we evaluate the performance of the particle filter method for model structure identification and calibration of a land use change CA. Furthermore, we assess the effect of the amount of observational data assimilated, because time series of good quality land use maps are often absent (Straatman et al., 2004). We also consider the effect of a pre-set (expert-based) model structure, to represent the situation of a model structure identification determined beforehand, which is now common practice in land use change modelling. In all approaches we provide confidence intervals with the land use projections, useful as a decision criterion for policy makers.

The assessments are carried out on a case study of the expansion of sugar cane fields in the São Paulo state in Brazil, using an adapted form of the PCRaster Land Use Change model (PLUC) (Verstegen et al., 2012). As the sugar cane is partly used to produce ethanol, this case study is relevant in view of the current debate on the sustainability of bioenergy from dedicated crops when land use change is taken into account (Lapola et al., 2010; Hellmann and Verburg, 2011). São Paulo is especially interesting because it has a long history in ethanol production (Walter et al., 2011) and very good observational data availability (Rudorff et al., 2010).

The next section provides a definition of the problem of transition rule identification in a CA, a brief explanation of data assimilation, a description of the case study, an outline of the prior information about the land use change model structure and parameters, details of the performance measures used, a description of the observational data, and a scenario sketch. This is followed by a combined results and discussion section, and a conclusion section.

2. Methods

2.1. Model structure and parameter identification in a land use change cellular automaton

A cellular automaton (CA) consists of a set of transition rules representing the processes that lead to change in the system state over time and rules to combine these transition rules (Fig. 1). In the case of a land use change CA a transition rule is a function calculating the suitability of each location (cell) for a particular land use type, with respect to a spatial attribute that influences the allocation of that land use type, for instance the slope or the distance to roads. So, a land use change CA contains for each land use type a set of transition rules. The transition rules contain parameters defining the characteristics of the process represented by the transition rule, for example an exponent in an exponential relationship between the suitability value and slope. The transition rules need to be selected and combined such that they represent the key processes that steer the spatial allocation of land use change. This can be accomplished by selecting from a set of candidate transition rules. This could be done either in a Boolean fashion, by switching transition rules on or off, or in a continuous fashion, by weighting each transition rule. We refer to this selection of transition rules as model structure identification.

In modelling, it is essential to find the model structure and parameter values that result in an optimal model representation of the studied land use system. Identification of the model structure and parameter values can be accomplished through comparison of the modelled system, with certain transition rules and

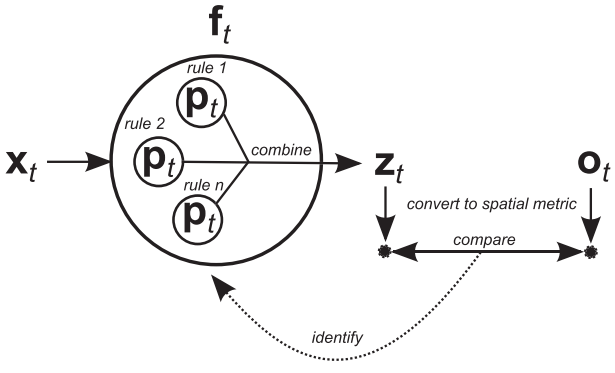

Fig. 1. Conceptual model of a general CA: represents the processes of change in the system state over time, i.e. the set of transition rules and the way to combine them, $\mathbf{x}_t$ represents all inputs, usually spatial attributes, and contains the parameters. Model calibration refers to identifying $\mathbf{p}_t$, model structure identification refers to selecting the transition rules. Identification of the parameters and model structure is based on a comparison between the land use map $\mathbf{z}_t$, or a derived spatial metric, with the observed land use map $\mathbf{o}_t$, or a derived spatial metric. The parameter values and model structure with the smallest difference between $\mathbf{z}_t$ and $\mathbf{o}_t$ are considered optimal.