Agriculture, Ecosystems and Environment 217 (2016) 99–110

Contents lists available at ScienceDirect

Agriculture, Ecosystems and Environment

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

Identification and apportionment of the drivers of land use change on a regional scale: Unbiased recursive partitioning-based stochastic model application

Qi Wang[1,a,*], Qingfu Ren[1,b,c], Jianfeng Liu[d]

[a] *Guangdong Key Laboratory of Agricultural Environment Pollution Integrated Control, Guangdong Institute of Eco-Environmental and Soil Sciences, Guangzhou, China*
[b] *College of Soil and Water Conservation, Beijing Forestry University, Beijing, China*
[c] *State Key Laboratory of Simulation and Regulation of Water Cycle in River Basin, China Institute of Water Resources and Hydropower Research, Beijing, China*
[d] *Research Institute of Forestry, Chinese Academy of Forestry, Beijing 100091, China*

ARTICLE INFO

Article history:
Received 27 July 2015
Received in revised form 1 November 2015
Accepted 2 November 2015

Keywords:
Land use change
Driver
Conditional inference tree
Random forest
Stochastic model
Regional scale

ABSTRACT

Land-system science places increasing emphasis on the complex, telecoupled and often nonlinear nature of land-system changes. Categorical diversity, spatial-temporal heterogeneity and the rapid change rate of land use drivers challenged researchers to determine this sophisticated nature of landscape changes at the regional scale. This study aims to develop a new framework, involving remote sensing, GIS and machine learning, to identify and apportion the important factors responsible for land use change at the regional scale. The Jiangxi province in China was used as a case study. The drivers of land use change were identified and apportioned using stochastic models based on unbiased recursive partitioning method embracing the conditional inference tree (CIT) and random forest (RF) with a focus on cropland and urban land. Regression trees for determining the major drivers of cropland and urban land change were established. Partial dependence plots and two-variable interaction plots from the resulting RF models for explaining the contributions of the drivers of cropland and urban land change were developed. A spatial autoregressive model was implemented as a supplement tool to help explain the causes of land use change. The determinants of cropland and urban land change were quantitatively assessed by CIT and RF along with the interactions of multiple drivers. The models were verified using rigorous out-of-sample and actual-versus-predicted testing. The results show strong suitability of unbiased recursive partitioning-based models to the assessment of the complex drivers of land use change at a regional scale

1. Introduction

Landscapes produce a wide range of valuable ecosystem services, not only for agricultural production but also for emissions and sequestration of greenhouse gases, open-access recreational visits, urban green space, and wild-species diversity (Bateman et al., 2013). Landscape heterogeneity and ecological processes affect the values from these services at the regional level, and current land use is vulnerable to the impacts of global change (Foley et al., 2005; Schroter et al., 2005). Land-system places increasing emphasis on the complex, telecoupled and often nonlinear nature of land-system changes. A comprehensive assessment of the drivers of land use associated with a systematic environmental and economic analysis can support cost-effective land planning and is of great importance (Turner et al., 1995; Lambin et al., 1999). Land use change is driven by multiple, interacting factors that originate from the local to the global scales; they form a complex system of dependencies, interactions and feedback loops (Lambin et al., 2003). Globally and historically, landscape changes appeared to track well with the population, affluence, technology and policy variables of the environmental impact because they captured the demand for land and resources together with the means by which they were supplied (Lambin

* Corresponding author. Fax: +86 20 87024766.
E-mail address: jenniferqi520@gmail.com (Q. Wang).
[1] These authors contributed equally to this work.

http://dx.doi.org/10.1016/j.agee.2015.11.004

et al., 2001; Turner et al., 2007). Categorical diversity, spatial-temporal heterogeneity and the rapid change rate of land use drivers challenged researchers to determine the causes of this sophisticated nature of landscape changes at the regional scale; exploring suitable methods to address the challenge is vital. To this end, models can serve as powerful tools for the identification and apportionment of drivers.

In modeling reviews, the categories of modeling approaches to land use change and drivers are summarised as follows: stochastic models, empirical-statistical models, process-based and mechanistic models, agent-based and economic models (Lambin et al., 2000). The common methods applying multivariate analyses for empirical predicting landscape change drivers are quantitative in nature and are unable to facilitate the inclusion of qualitative data. Further, due to the complexity of the interactions of drivers that lead to land use change, there is often an understanding of the relationship within the process, but modeling is generally constrained by the absence of consistent and comprehensive knowledge linking to the process (Parker et al., 2008). The modeling of spatial dependency can use the interactions of drivers across space including transition patterns, migration, economy and commodity flows (Aspinall, 1993). Although the change process is often not the same within a region, the direction of change may be navigated by dominant drivers or constraints such as policy, planning or economic rewards. The agent-based models in the study of coupled human-land systems provide one means of representing behaviors of individual and interacting land managers while taking account of cognitive, social, economic and environmental constraints (Hare and Deadman, 2004; Aspinall, 2009; Gotts and Polhill, 2009). For example, Bayesian belief networks (BBN) provide a means to utilise a combination of limited empirical data and expert knowledge (Aalders and Aitkenhead, 2006; Marcot et al., 2006). While the use of expert knowledge to build a model is often criticised, developing the BBN in an evolving knowledge development process can be very useful in the early phases of exploring complex processes. Modeling approaches are best suited to take account of past location, rate of change and drivers. Stochastic models show superiority to understand the complex relationship between land use changes and their drivers, and current stochastic models in land use research are mainly based on the transition probability method (Soares et al., 2002). The unbiased recursive partitioning method provided in this study offers a new opportunity to explore the spatial implications of land use drivers at the regional scale. The rationale of the method is a unified framework embedding recursive binary partitioning with piecewise constant fits into the well-defined theory of permutation tests. The conditional distribution of statistics testing the correlation between responses and predictors is the basis for an unbiased selection among predictors measured at different scales. Moreover, multiple test procedures are used to determine whether a significant correlation exits between any of the predictors and stopping criteria (Hothorn et al., 2006). Such statistically motivated stopping criteria lead to the predictive performance of regression models is equivalent to that of optimally pruned trees, offering a computationally efficient solution to overfitting. We applied unbiased recursive partitioning-based stochastic models, the conditional inference tree (CIT) and the random forest (RF), which are capable of handling spatial autocorrelation, extracting the importance of driving forces and gaining their interactions, to identify and apportion land use drivers on a regional scale.

Compared to empirical statistical models such as linear regression and discriminant analysis, recursive partitioning method shows dominant advantages, which can address different types of response variables, make faithful data descriptions without strong model assumptions, handle a large number of both quantitative and qualitative input variables, and easily interpret the results (Lee et al., 2006). Classic decision trees such as the classification and regression tree (CART) (Breiman et al., 1984) and the "C5" (Quinlan, 1993) have two major limitations: over-fitting and biased variable selection (Bradford et al., 1998). These trees are developed by selecting the best split measured by the Gini index (Breiman et al., 1984) or the information obtain through an exhaustive search over all possible variables to split and all possible places for a split. Although pruning based on cross-validation can be used to avoid the over-fitting, the biased variable selection significantly complicates the interpretation of the decision trees (Hothorn et al., 2006; Nagy et al., 2010). CIT can handle different types and scales of the explanatory variables against over-fitting and biased variable selection (Hothorn et al.,

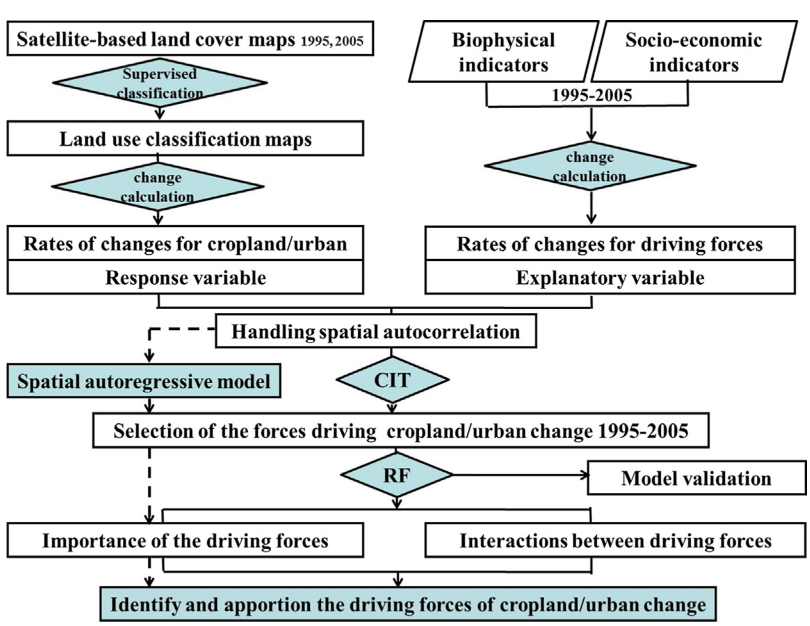

Fig. 1. Unbiased recursive partitioning-based method for identifying and apportioning the drivers of cropland and urban land change.