Geosci. Model Dev. Discuss., 8, 3359–3402, 2015
www.geosci-model-dev-discuss.net/8/3359/2015/
doi:10.5194/gmdd-8-3359-2015

Geoscientific Model Development Discussions
Open Access

This discussion paper is/has been under review for the journal Geoscientific Model Development (GMD). Please refer to the corresponding final paper in GMD if available.

An open and extensible framework for spatially explicit land use change modelling in R: the lulccR package (0.1.0)

S. Moulds[1,2], W. Buytaert[1,2], and A. Mijic[1]

[1]Department of Civil and Environmental Engineering, Imperial College London, London, UK
[2]Grantham Institute for Climate Change, Imperial College London, London, UK

Received: 12 March 2015 – Accepted: 23 March 2015 – Published: 29 April 2015

Correspondence to: S. Moulds (simon.moulds10@imperial.ac.uk)

Published by Copernicus Publications on behalf of the European Geosciences Union.

Abstract

Land use change has important consequences for biodiversity and the sustainability of ecosystem services, as well as for global environmental change. Spatially explicit land use change models improve our understanding of the processes driving change and make predictions about the quantity and location of future and past change. Here we present the lulccR package, an object-oriented framework for land use change modelling written in the R programming language. The contribution of the work is to resolve the following limitations associated with the current land use change modelling paradigm: (1) the source code for model implementations is frequently unavailable, severely compromising the reproducibility of scientific results and making it impossible for members of the community to improve or adapt models for their own purposes; (2) ensemble experiments to capture model structural uncertainty are difficult because of fundamental differences between implementations of different models; (3) different aspects of the modelling procedure must be performed in different environments because existing applications usually only perform the spatial allocation of change. The package includes a stochastic ordered allocation procedure as well as an implementation of the widely used CLUE-S algorithm. We demonstrate its functionality by simulating land use change at the Plum Island Ecosystems site, using a dataset included with the package. It is envisaged that lulccR will enable future model development and comparison within an open environment.

1 Introduction

Land use and land cover change is degrading biodiversity worldwide and threatening the sustainability of ecosystem services upon which individuals and communities depend (Turner et al., 2007). Cumulatively, it is a major driver of global and regional environmental change (Foley, 2005). For example, as a result of extensive deforestation in Central and South America and Southeast Asia land use and land cover change is the second largest anthropogenic source of carbon dioxide (Le Quéré et al., 2009), while the conversion of rainfed agriculture and natural land cover to intensively managed agricultural systems in northwest India is now putting severe pressure on regional water resources (Rodell et al., 2009; Shankar et al., 2011; Wada et al., 2012). In addition, land use and land cover change may influence local and regional climate through its impact on the surface energy and water balance (Pitman et al., 2009; Seneviratne et al., 2010; Boysen et al., 2014). Land use change models are widely used to understand and quantify key processes that affect land use and land cover change and simulate past and future change under different scenarios and at different spatial scales (Veldkamp and Lambin, 2001; Mas et al., 2014). The output of these models may be used to support decisions about local and regional land use planning and environmental management (e.g. Couclelis, 2005; Verburg and Overmars, 2009) or investigate the impact of change on biodiversity (e.g. Nelson et al., 2010; Rosa et al., 2013), water resources (e.g. Li et al., 2007; Lin et al., 2008; Rodríguez Eraso et al., 2013) and climate variability (e.g. Sohl et al., 2007, 2012).

Land use and land cover change is the result of complex interactions between different biophysical and socioeconomic conditions that vary across space and time (Verburg et al., 2002; Overmars et al., 2007). Several different model structures have been devised to capture this complexity and meet different objectives. Some models operate at the global or regional scale to estimate the quantity of land use change at national or subnational levels based on economic considerations (e.g. Souty et al., 2012), whereas spatially explicit models, the focus of the present study, operate over a spatial grid to predict the location of land use change (Mas et al., 2014). Inductive spatially explicit models are based on predictive models that predict the suitability of each model grid cell as a function of spatially explicit predictor variables, while deductive models predict the location of change according to specific theories about the processes driving change (Overmars et al., 2007; Magliocca and Ellis, 2013). Inductive and deductive models operating at different spatial scales may be combined to better represent the complexity of a system (e.g. Castella and Verburg, 2007; Moreira et al., 2009). The

main output of land use change models is a set of land use maps depicting the location of change over time. Detailed reviews of different models and modelling approaches are available in Verburg et al. (2004), Brown et al. (2013) and Mas et al. (2014).

Spatially explicit land use change models are commonly written in compiled languages such as C/C++ and Fortran and distributed as software packages or extensions to proprietary geographic information systems such as ArcGIS or IDRISI. As Rosa et al. (2014) points out, it is uncommon for the source code of model implementations to be made available (e.g. Verburg et al., 2002; Soares-Filho et al., 2002; Verburg and Overmars, 2009; Schaldach et al., 2011). While it is true that the concepts and algorithms implemented by the software are normally described in scientific journal articles, this fails to ensure the reproducibility of scientific results (Peng, 2011; Morin et al., 2012), even in the hypothetical case of a perfectly described model (Ince et al., 2012). In addition, running binary versions of software makes it difficult to detect silent faults (faults that change the model output without obvious signals), whereas these are more likely to be identified if the source code is open (Cai et al., 2012). Moreover, it forces duplication of work and makes it difficult for members of the scientific community to improve the code or adapt it for their own purposes (Morin et al., 2012; Pebesma et al., 2012; Steiniger and Hunter, 2013).

Current software packages usually exist as specialised applications that implement one algorithm. Indeed, it is common for applications to perform only one part of the modelling process. For example, the Change in Land Use and it Effects at Small regional extent (CLUE-S) software only performs spatial allocation, requiring the user to prepare model input and conduct the statistical analysis upon which the allocation procedure depends elsewhere (Verburg et al., 2002). This is time consuming and increases the likelihood of user errors because inputs to the various modelling stages must be transferred manually between applications. Furthermore, very few programs include methods to validate model output, which could be one reason for the lack of proper validation of models in the literature, as noted by Rosa et al. (2014). The lack of a common interface amongst land use change models is problematic for the community because there is widespread uncertainty about the appropriate model form and structure for different applications (Verburg et al., 2013). Under these circumstances it is useful to experiment with different models to identify the model that performs best in terms of calibration and validation (Schmitz et al., 2009). Alternatatively, ensemble modelling may be used to understand the impact of structural uncertainty on model outcomes (Knutti and Sedláček, 2012). This approach has been used successfully in the CMIP5 experiments (Taylor et al., 2012; Knutti and Sedláček, 2012), global and regional drought prediction (Tallaksen and Stahl, 2014; Prudhomme et al., 2014) and species distribution modelling (Grenouillet et al., 2011), for example. However, while some land use change model comparison studies have been carried out (e.g Pérez-Vega et al., 2012; Mas et al., 2014; Rosa et al., 2014), fundamental differences between models in terms of scale, resolution and model inputs prevent the widespread use of ensemble land use change predictions (Rosa et al., 2014). As a result, the uncertainty associated with model outcomes are rarely communicated in a formal way, raising questions about the utility of such models (Pontius and Spencer, 2005).

An alternative approach is to develop frameworks that allow several different modelling approaches to be implemented within the same environment. One such application is the PCRaster software, a free and open source GIS that includes additional capabilities for spatially explicit dynamic modelling (Schmitz et al., 2009). The PCRcalc scripting language and development environment allows users to build models with native PCRaster operations such as map algebra and neighbourhood functions. Alternatively, the PCRaster application programming interface (API) allows users to extend the functionality of PCRaster in different programming languages using native and external data types (Schmitz et al., 2009). For example, the current version of FALLOW (van Noordwijk, 2002; Mulia et al., 2014), a deductive model that simulates farmer decisions about agricultural land use in response to biophysical and socioeconomic driving factors, is built using the PCRaster framework. TerraME (Carneiro et al., 2013) is a platform to develop models for simulating interactions between society and the environment. It provides more flexibility than PCRaster because models can be composed

of coupled sub-models with different temporal and spatial resolutions (Moreira et al., 2009; Carneiro et al., 2013). The platform is built on the open source TerraLib geospatial library (Câmara et al., 2008) which handles different spatio-temporal data types, includes an API for coupling the library with R (R Core Team, 2014) to perform spatial statistics, and supports dynamic modelling with cellular automata. The LuccME extension to TerraME includes current implementations of CLUE and CLUE-S (Veldkamp and Fresco, 1996; Verburg et al., 1999), an earlier version of CLUE-S that operates at larger spatial scales, written in Lua.

The R environment is a free and open source implementation of the S programming language, a language designed for programming with data (Chambers, 2008). Although the development of R is strongly rooted in statistical software and data analysis, it is increasingly used for dynamic simulation modelling in diverse fields (Petzoldt and Rinke, 2007). Additionally, in the last decade it has become widely used by the spatial analysis community, largely due to the *sp* package (Pebesma and Bivand, 2005; Bivand et al., 2013) which unified many different approaches for dealing with spatial data in R and allowed subsequent package developers to use a common framework for spatial analysis. The *rgdal* package (Bivand et al., 2014) allows R to read and write formats supported by the Geospatial Data Abstraction Library (GDAL) and OGR library. Through the *raster* package (Hijmans, 2014), R now includes most functions for raster data manipulation commonly associated with GIS software. Building on these capabilities, several R packages have been created for dynamic, spatially explicit ecological modelling (e.g. Petzoldt and Rinke, 2007; Fiske and Chandler, 2011). In addition, two recent land use change models have been written for the R environment. StocModLCC (Rosa et al., 2013) is a stochastic inductive land use change model for tropical deforestation while SIMLANDER (Hewitt et al., 2013) is a stochastic cellular automata model to simulate urbanisation. Thus, R is well suited for spatially explicit land use change modelling. To date, however, R has not been used to develop a framework for land use change model development and comparison. In this paper we describe the *lulccR* package, a free and open source software package for land use change modelling in the R environment.

2 Design goals

The first design goal of *lulccR* is to provide a framework that allows users to perform every stage of the modelling process, shown by Fig. 1, within the same environment. It therefore includes methods to process and explore model input, fit and evaluate predictive models, estimate the fraction of the study area belonging to each land use type at different timesteps, allocate land use change spatially, validate the model and visualise model outputs. This provides many advantages over specialised software applications. Firstly, it improves efficiency and reduces the likelihood of user errors because intermediate inputs and outputs exist in the same environment (Fiske and Chandler, 2011; Pebesma et al., 2012). Secondly, it encourages interactive model building because different aspects of the procedure can easily be revisited. Thirdly, it means it is straightforward to investigate the effect of different inputs and model setups on model outcomes. Finally, and perhaps most importantly, it improves the reproducibility of scientific results because the entire modelling process can be expressed programmatically (Pebesma et al., 2012).

lulccR is intended as an alternative to current paradigm of closed-source, specialised software packages which, in our view, disrupt the scientific process. Thus, the second design goal is to create an open and extensible framework allowing users to examine the source code, modify it for their own purposes and freely distribute changes to the wider community. The package exploits the openness of the R system, particularly with respect to the package system, which allows developers to contribute code, documentation and data sets in a standardised format to repositories such as the Comprehensive R Archive Network (CRAN) (Pebesma et al., 2012; Claes et al., 2014). As a result of this philosophy R users have access to a wide range of sophisticated tools for data management, spatial analysis and plotting and visualisation. Significantly, given

the importance of predictive models to land use change modelling, R has become the standard development platform for statistical software.

One of the consequences of providing a modelling framework in R is that users of the software must become programmers (Chambers, 2000). We recognise that this represents a different approach to the current practice of providing land use change software packages with graphical user interfaces (GUIs), and acknowledge that for users unfamiliar with programming it could present a steep learning curve. Therefore, the third design goal is to provide software that is easy to use and accessible for a users with different levels of programming experience. The package includes complete working examples to allow beginners to start using the package immediately from the R command shell, while more advanced users should be able to develop modelling applications as scripts. Furthermore, the package is designed to be extensible so that users can contribute new or existing methods. Similarly, the source code of *lulccR* is accessible so that users can locate the methods in use and understand algorithm implementations. Acknowledging that many scientists lack any formal training in programming (Joppa et al., 2013; Wilson et al., 2014), we hope this final goal will ensure the software is useful for educational purposes as well as scientific research.

3 Software description

To achieve the design goals we adopted an object-oriented approach. This provides a formal structure for the modelling framework which allows the different stages of land use change modelling applications to be handled efficiently. Furthermore, it encourages the reuse of code because objects can be used multiple times within the same application or across different applications. It is extensible because it is straightforward to extend existing classes using the concept of inheritance or create new methods for existing classes. In *lulccR* we use the S4 class system (Chambers, 1998, 2008), which requires classes and methods to be formally defined. This system is more rigorous than the alternative S3 system because objects are validated against the class definition when they are created, ensuring that objects behave consistently when they are passed to functions and methods. Figure 2 shows the class diagram for *lulccR* together with a list of the most important functions. Here we describe the main components of *lulccR* integrated with an example application for the Plum Island Ecosystems dataset to demonstrate its functionality.

3.1 Data

The failure to provide driving data for land use change modelling exercises alongside published literature is identified by Rosa et al. (2014) as a major weakness of the discipline. The *lulccR* package includes two datasets that have been widely used in the land use change community, allowing users to quickly start exploring the modelling framework.

3.1.1 Plum Island Ecosystems

The Plum Island Ecosystems Long Term Ecological Research site is located in northeast Massachusetts and includes the watersheds of the Ipswich River, Parker River and Rowley River (http://pie-lter.ecosystems.mbl.edu/). Research at the site aims to understand the response of coastal ecosystems to changes in land use, climate and sea level (Hobbie et al., 2003; Alber et al., 2013). In recent decades the area, which is located approximately 50 km from the Boston, has undergone extensive land use change from forest to residential use (Aldwaik and Pontius, 2012). This has altered the hydrological behaviour of the three watersheds with negative impacts on downstream ecosystems (Morse and Wollheim, 2014). The dataset included in *lulccR* was originally developed as part of the MassGIS program (MassGIS, 2015) but has been processed by Pontius and Parmentier (2014). Land use maps depicting forest, residential and other uses are available for 1985, 1991 and 1999. Although MassGIS provides a fourth land use map for 2005 this was produced using a different classification methodology and cannot be used for change detection (Morse and Wollheim, 2014). Three predictor

variables are included: elevation, slope and distance to built land in 1985. Land use for the site in 1985 is shown by Fig. 3.

3.1.2 Sibuyan

Sibuyan is a small island with a total area of 456 km^2 belonging to Romblon province in the Phillipines. The central region is mountainous and heavily forested while the surrounding area is used for natural land cover, plantations, agriculture and other uses (Verburg et al., 2002). The island is relevant for land use change studies because its rich biodiversity is threatened by illegal logging and unsustainable farming practices (Villamor and Lasco, 2009). The dataset included in *lulccR* is an adapted version of the dataset distributed with the CLUE-S model, and includes includes an observed land use map for 1997, a number of predictor variables, a map of the Mount Guiting-Guiting Natural Park, a protected area in the centre of the island, and four demand scenarios for the period 1997 to 2011. In addition, we include the simulated map for 2011 from the original CLUE-S software, corresponding to the first demand scenario, for benchmarking purposes. The naming convention of this map follows that of the observed land use map for 1997. Further information about Sibuyan island in the context of land use change is provided elsewhere in Verburg et al. (2002, 2004), for example.

3.2 Data processing

One of the most challenging aspects of land use change modelling is to obtain and process the correct input data. In *lulccR* all spatially explicit input data must be stored in one of the file types supported by *rgdal* or exist in the R workspace as a raster object belonging to the *raster* package (RasterLayer, RasterStack or RasterBrick). The most fundamental input required by land use change models is an initial map of observed land use, which is typically obtained from classified remotely sensed data. This map represents the initial condition for model simulations and, for inductive modelling, it is used to fit predictive models. Sometimes it is more useful to consider observed land use transitions: in this case an additional map for an earlier timestep is required, as shown by Fig. 1. Ideally, two more observed land use maps for subsequent timesteps should be obtained for calibrating and validating the land use change model (Pontius et al., 2004a).

In *lulccR* observed land use data are represented by the `ObsLulcMaps` class. In the following code snippet we install the *lulccR* package from github and create an `ObsLulcMaps` object for the Plum Island Ecosystem dataset:

```
> library(devtools)
> install_github("simonmoulds/r_lulccR", subdir="lulccR")
> library("lulccR")
> obs <- ObsLulcMaps(x=pie,
                     pattern="lu",
                     categories=c(1,2,3),
                     labels=c("Forest","Built","Other"),
                     t=c(0,6,14))
```

The `ObsLulcMaps` object is important to land use change studies because it defines the spatial and temporal domain of subsequent operations. Thus, additional spatial data to be included in the model input should have the same characteristics as the maps contained in the corresponding `ObsLulcMaps` object (however, some helper functions are available to resample maps to the correct spatial resolution). The `t` argument in the constructor function specifies the timesteps associated with the observed land use maps. The first timestep must always be zero; if additional maps are present they should have timesteps greater than zero, even in backcast models. In most land use change modelling applications a timestep represents one year but there is no requirement for this to be the case.

A useful starting point in land use change modelling is to obtain a transition matrix for two observed land use maps from different times to identify the main historical transitions in the study region (Pontius et al., 2004b), which can be used as the basis for fur-

ther research into the processes driving change. In *lulccR* we use the `crossTabulate` function for this purpose:

```
> crossTabulate(x=obs, index=c(1,2))
```

The output of this command reveals that for the Plum Island Ecosystems site the dominant change between 1985 and 1999 was the conversion of forest to built areas.

Inductive and deductive land use change models predict the location of change based on spatially explicit biophysical and socioeconomic explanatory variables. These may be static, such as elevation or geology, or dynamic, such as maps of population density or road networks. In *lulccR* these two types of explanatory variable are separated by a simple naming convention, which is explained in detail in the package documentation (see Supplement). Collectively, they are represented by an object of class `ExpVarMaps`, which can be created as follows:

```
> ef <- ExpVarMaps(x=pie, pattern="ef")
```

Apart from observed land use maps and predictor variables other input maps may be required. The two allocation routines currently included with *lulccR* accept a mask file, which is used to prevent change within a certain geographic area such as a national park or other protected area, and a land use history file, which is used as the basis for certain decision rules. These are handled by *lulccR* as standard RasterLayer objects.

3.3 Predictive modelling

Inductive land use change models are based on predictive models which relate the pattern of observed land use to spatially explicit explanatory variables. Logistic regression is the most widely used model type (e.g. Pontius and Schneider, 2001; Verburg et al., 2002), however, there is growing interest in the application of local and non-parametric models to inductive land use change modelling (e.g. Tayyebi et al., 2014). Currently *lulccR* supports binary logistic regression, available in base R, recursive partitioning and regression trees, provided by the *rpart* package (Therneau et al., 2014), and random forests, provided by the *randomForest* package (Liaw and Wiener, 2002). In all cases a separate model must be obtained for each land use type in the study region. *lulccR* does not provide additional functionality to fit predictive models to the observed data since R is already optimised for this purpose.

Parametric models such as logistic regression models assume the input data to be independent and identically distributed (Overmars et al., 2003; Wu et al., 2009). In spatial analysis this assumption is often violated because of spatial autocorrelation, which reduces the information content of an observation because its value can to some extent be predicted by the value of its neighbours (Beale et al., 2010). While non-parametric models such as regression trees and random forest make no assumption of independence, a recent study by Mascaro et al. (2014) showed that these models may nevertheless be affected by spatial autocorrelation. Dormann et al. (2007) discusses several ways to account for spatial autocorrelation, however, the simplest, and most widely used, approach is to fit the models to a random subset of the data (e.g. Verburg et al., 2002; Wassenaar et al., 2007; Echeverria et al., 2008). This method is provided in *lulccR* using the `createDataPartition` function of the *caret* package (Kuhn et al., 2012) to perform a stratified random sample of the data. The data partition is obtained as follows:

```
> part <- partition(x=obs@maps[[1]], size=0.5, spatial=FALSE)
```

returning a named list object with the index of cells in the three partitions (training, testing, all cells). To fit models in R it is necessary to supply a formula and a data frame (the main data structure in R) containing the response and explanatory variables. The predictive models we use aim to predict the presence or abscence of each land use type; thus, it is first necessary to convert the observed land use maps to binary response variables before fitting a model to each land use. A typical workflow is shown here:

```
> br <- raster::layerize(obs@maps[[1]])
> names(br) <- obs@labels
> train.df <- raster::extract(x=br, y=part$train, df=TRUE)
```



```
> train.df <- cbind(train.df, as.data.frame(x=ef, cells=part$train))
> built.glm <- glm(Built~ef_001+ef\_002+ef_003,
                   family=binomial,
                   data=train.df)
```

where the final command fits a binary logistic regression model to predict the occurrence of built based on three explanatory variables (elevation, slope and distance to 1971 built area). This procedure is repeated for each land use in the study area, which, for the Plum Island Ecosystems dataset, includes forest and other land uses in addition to built. For forest we employ a null model (a model with no explanatory factors) because the transition from forest to built is determined by the location suitability of built rather than that of forest. Of the predictive models supported by *lulccR* only binary logistic regression permits a null model to be fitted. Predictive models for each land use are represented by an object of class `PredModels`:

```
> glm.models <- PredModels(list(forest.glm, built.glm, other.glm),
                           obs=obs)
```

The resulting object makes it straightforward to plot the suitability of each land use over the study region using the `calcProb` function in combination with some additional functionality from the *raster* package (see Supplement). The resulting plot is shown by Fig. 4.

Methods to evaluate statistical models are provided by the *ROCR* package (Sing et al., 2005), allowing the user to assess model performance using several methods including the receiver operator characteristic (ROC), which is widely used to measure the performance of models predicting the presence or abscence of a phenomenon. This method uses a threshold to transform an index variable, in our case the output of the predictive models which varies between zero and one, to a boolean variable where values above the threshold are true (1) and values below the threshold are false (0). The transformed variable is compared to reference information to generate a contingency table with entries for true positives, false positives, true negatives and false negatives. The ROC considers multiple thresholds in order to plot a curve of true positive rate against false positive rate (Pontius and Parmentier, 2014). It is often summarised by the area under the curve (AUC), where one indicates a perfect fit and 0.5 indicates a purely random model.

In *lulccR* we extend the native *ROCR* classes to better suit our purposes. The `prediction` method of *ROCR* is extended by `Prediction` to handle one or more `PredModels` objects to enable comparison of different types of predictive model. The resulting object is then used to create a `Performance` object, for which a plot method exists. The procedure to evaluate several `PredModels` objects is as follows:

```
> glm.pred <- Prediction(models=glm.models,
                         obs=obs,
                         ef=ef,
                         partition=part$test)
> glm.perf <- Performance(pred=glm.pred, measure="rch")

# not shown: create rpart.perf and rf.perf in the same way
> p <- Performance.plot(list(glm=glm.perf,
                        rpart=rpart.perf,
                        rf=rf.perf))
> print(p) # ROC plots
```

Figure 5 shows the ROC plots for each land use type and for each type of predictive model supported by *lulccR*. The plots show that binary logistic regression and random forest models perform similarly for built and other land uses, while regression tree models perform least well in both cases.

3.4 Demand

Spatially explicit land use change models are normally driven by non-spatial estimates of the total area occupied by each land use type at every timestep. This means regional drivers of land use change, such as population growth and technology, are considered implicitly (Fuchs et al., 2013). While some models calculate demand at each timestep based on the spatial configuration of the landscape at the previous timestep (e.g. Rosa et al., 2013), it is more common to supply land use area for every timestep at the

beginning of the simulation (e.g. Pontius and Schneider, 2001; Verburg et al., 2002; Sohl et al., 2007). This is the approach currently supported in *lulccR*. Land use area may be estimated using non-spatial land use models or, if the study aims to reconstruct historic land use change, national and subnational land use statistics may be used (e.g. Ray and Pijanowski, 2010; Fuchs et al., 2013). *lulccR* includes a function to interpolate or extrapolate land use area based on two or more observed land use maps: this approach is often used to predict the quantity of land use change in the near-term (Mas et al., 2014). For the current example we obtain land use demand for each year between 1985 and 1999 by linear interpolation, as follows:

```
> dmd <- approxExtrapDemand(obs=obs, tout=0:14)
```

In reality we are not usually interested in simulating land use change between two known points. However, doing so is useful for model validation: we test the ability of the model to predict the location of change given the exact quantity of change.

3.5 Allocation

The allocation component of land use change models estimates the location of change in the study region at each timestep (Verburg et al., 2002). Currently *lulccR* includes two inductive allocation routines: an implementation of the CLUE-S algorithm and a stochastic ordered procedure based on the algorithm described by Fuchs et al. (2013). Before running either allocation procedure *lulccR* implements a number of decision rules to identify the specific land use transitions that are allowed at each location. While the set of rules included in *lulccR* could be used as the basis of a simple agent-based model of the type employed by Castella and Verburg (2007), their main purpose is to allow the modeller to include additional knowledge about the system while still relying on an inductive allocation procedure. The container class `ModelInput` represents the different inputs to the allocation function and checks that the objects are compatible with each other:

Discussion Paper | Discussion Paper | Discussion Paper | Discussion Paper |

```
> input <- ModelInput(obs=obs,
                      ef=ef,
                      models=glm.models,
                      time=0:14,
                      demand=dmd)
```

These objects are supplied as the main input to objects inheriting from the virtual class `Model`, which represents standard information required by the two allocation routines currently implemented in *lulccR* and, indeed, most allocation routines described in the literature. Subclasses of `Model` are associated with a particular allocation method. These classes inherit general information held in `Model` and include specific information such as parameters and additional spatial input such as mask and land use history files. A generic `allocate` function receives objects inheriting from class `Model` and performs the relevant allocation routine. All methods belonging to the generic `allocate` function update the `Model` object with the allocation results. This design ensures that it is easy to add additional allocation routines to *lulccR*: developers simply need to define a new subclass of `Model` and write a new `allocate` method. Here we describe the decision rules and allocation routines currently available in *lulccR*.

3.5.1 Decision rules

The first decision rule included in *lulccR* is used to prohibit certain land use transitions. For example, in most situations it is unlikely that urban areas will be converted to agricultural land because the initial cost of urban development is high (Verburg et al., 2002). The second rule specifies a minimum number of timesteps before a certain transition is allowed, while the third rule specifies a maximum number of timesteps after which change is not allowed. These rules are used to control land use transitions that are time-dependent. For example, the transition from shrubland to closed forest is slow and cannot occur after only one year (Verburg and Overmars, 2009), whereas for some types of agriculture a location is only suitable for a certain number of growing seasons

because of declining soil quality. The fourth rule prohibits transitions to a certain land use in cells that are not within a user-defined neighbourhood of cells already belonging to the same land use. This rule is particularly relevant to cases of deforestation or urbanisation because this sort of change usually occurs at the boundaries of existing forests or cities, respectively.

Within the `allocate` function the first four decision rules are implemented by the `allow` function while the fifth decision rule is performed by the `allowNeighb` function. To apply neighbourhood rules it is necessary to supply corresponding neighbourhood maps to the allocation routine. In *lulccR* these are represented by the `NeighbMaps` class. Objects of this class are created with the following command:

```
> nb <- NeighbMaps(x=obs@maps[[1]],
                   categories=2,
                   weights=3)
```

Essentially, the `allow` and `allowNeighb` functions identify disallowed transitions according to the decision rules and set the suitability of these cells to NA. These transitions are ignored by the allocation routine. Care should be taken to ensure that after any decision rules are taken into account there are sufficient cells eligible to change in order to meet the specified demand at each timestep.

3.5.2 CLUE-S allocation method

The CLUE-S model implements an iterative procedure to meet the specified demand at each timestep. The model is summarised briefly here: for a full description see Verburg et al. (2002) and Castella and Verburg (2007). In the first instance each cell is allocated to the land use with the highest suitability as determined by the predictive models. Whereas the original CLUE-S model is based on binary logistic regression, *lulccR* allows any predictive model supported by `PredModels` to be used. After this step the suitability is increased for land uses where the allocated area is less than demand and decreased for land uses where it is greater than demand. The extent to

Discussion Paper | Discussion Paper | Discussion Paper | Discussion Paper |

which the suitability is increased or decreased is a function of the difference between allocated change and demand. This procedure is repeated until the demand is met. The original model perturbs the location suitability to limit the influence of nominal differences in land use suitability on the final model solution. This is replicated in *lulccR* except the user has greater control over the degree of perturbation. In effect, therefore, this parameter can be used to make the procedure more or less stochastic. In the following code snippet we first set the decision rules to allow all possible transitions and then define some parameter values. Then, we create an object of class `CluesModel` and pass this to the `allocate` function:

```
> clues.rules <- matrix(data=c(1,1,1,1,1,1,1,1,1),
                        nrow=3, ncol=3, byrow=TRUE)
> clues.parms <- list(jitter.f=0.0002,
                      scale.f=0.000001,
                      max.iter=1000,
                      max.diff=50,
                      ave.diff=50)
> clues.model <- CluesModel(x=input,
                            elas=c(0.2,0.2,0.2),
                            rules=clues.rules,
                            params=clues.parms)
> clues.model <- allocate(clues.model)
```

As an iterative procedure the CLUE-S algorithm employs for-loops, which are slow in R. To overcome this limitation we have written the CLUE-S procedure as a C extension using the .Call interface. To benchmark our version of CLUE-S we compared our simulated land use map for Sibuyan Island for 1997 with that of the original model using comparable model inputs. The results of the comparison, shown by Fig. 6, demonstrate that, while the model versions do not perform identically, the model results are certainly comparable. Due to limitations of the original model interface we couldn't use

this model to simulate land use change for the Plum Island Ecosystems dataset and therefore further verification was not possible.

3.5.3 Ordered method

The ordered allocation method is based on the algorithm described by Fuchs et al. (2013). The approach is less computationally expensive and more stable than the CLUE-S implementation because it does not simulate competition between different land use types. Instead, land allocation is performed in a hierarchical way according to the perceived socioeconomic value of each land use type. For land uses with increasing demand only cells belonging to land uses with lower socioeconomic value are considered for conversion. In this case, *n* cells with the highest location suitability for the current land use are selected for change, where *n* equals the number of transitions required to meet the demand. The converted cells, as well as the cells that remain under the current land use, are masked from subsequent operations. For land uses with decreasing demand only cells belonging to the current land use are allowed to change. Here, *n* cells with the lowest location suitability are converted to a temporary class which can be allocated to subsequent land uses. The land use with lowest socioeconomic value is a special case because it is considered last and, therefore, the number of cells that have not been assigned to other land uses must equal the demand for this land use. In practice, this means that the location suitability for this class has no influence on the result. We modify the algorithm described by (Fuchs et al., 2013) to allow stochastic transitions. If this option is selected, the location suitability of each cell allowed to change is compared to a random number between zero and one drawn from a uniform distribution. If demand for the land use is increasing only cells where the location suitability is greater than the random number are allowed to change, whereas for decreasing demand only cells where it is less than the random number are allowed to change.

In *lulccR* the ordered model is represented by the `OrderedModel` class. In the following code we create an `OrderedModel` object, supplying the order in which to allocate change (built, forest, other), and pass this to the generic `allocate` function:

```
> ordered.model <- OrderedModel(x=input,
                                order=c(3,1,2))
> ordered.model <- allocate(ordered.model)
```

3.6 Validation

Spatially explicit land use change models are validated by comparing the initial observed map with an observed and simulated map for a subsequent timestep (Pontius et al., 2011). Previous studies have extracted useful information from the three possible two-map comparisons (e.g. Pontius et al., 2007), however, recently Pontius et al. (2011) devised the concept of a three-dimensional contingency table to compare the three maps simulataneously. Not only is this approach more parsimonious, it also yields more information about allocation performance (Pontius et al., 2011). For example, from the table it is straightforward to identify different sources of agreement and disagreement considering all land use transitions, all transitions from one land use or a specific transition from one land use to another. In addition, it is possible to separate agreement between maps due to persistence from agreement due to correctly simulated change. This is important because in most applications the quantity of change is small compared to the overall study area (Pontius et al., 2004b; van Vliet et al., 2011), giving a high rate of total agreement which can misrepresent the actual model performance. It is common to perform the validation procedure at multiple resolutions because comparison at the native resolution of the three maps fails to separate minor allocation disagreement, which refers to allocation disagreement at the native resolution that is counted as agreement at a coarser resolution, and major allocation disagreement, which refers to allocation disagreement at the native resolution and the coarse resolution (Pontius et al., 2011).

In *lulccR*, three-dimensional contingency tables at different resolutions are represented by the `ThreeMapComparison` class. Two subclasses of `ThreeMapComparison` represent different types of information that can be extracted from the tables: the `AgreementBudget` class represents sources of agreement and disagreement between the three maps at different resolutions while the `FigureOfMerit` class represents figure of merit scores. This measure, which is useful to summarise model performance, is defined as the intersection of observed and simulated change divided by the union of these (Pontius et al., 2011), such that a score of one indicates perfect agreement and a score of zero indicates no agreement. Plotting functions for `AgreementBudget` and `FigureOfMerit` objects allow the user to visualise model performance at different resolutions. The ordered model output for Plum Island Ecosystems is validated in the following way:

```
> ordered.tabs <- ThreeMapComparison(x=ordered.model,
                                     timestep=14,
                                     factors=2^(1:10))
> ordered.agr <- AgreementBudget(x=ordered.tabs)
> ordered.agr.p <- AgreementBudget.plot(ordered.agr, from=1, to=2)
> ordered.fom <- FigureOfMerit(x=ordered.tabs)
> ordered.fom.p <- FigureOfMerit.plot(ordered.fom, from=1, to=2)
```

This procedure was repeated for the CLUE-S model output. The agreement budgets for the transition from forest to built for the two model outputs are shown by Fig. 7, while Fig. 8 shows the corresponding figure of merit scores.

4 Discussion

The example application for Plum Island Ecosystems demonstrates the key strengths of the *lulccR* package. Firstly, it allows the entire modelling procedure to be carried out in the same environment, reducing the likelihood of mistakes that commonly arise when data and models are transferred between different software packages. A framework in R specifically allows users to take advantage of a wide range of statistical and machine learning techniques for predictive modelling, and, because R is widely regarded as the de facto standard for statistical model development, users of the package will have access to the most recent developments in these fields. The framework allows users to experiment with different model structures interactively and provides methods to quickly compare different model outputs. The example also highlights the advantages of an object-oriented approach: land use change modelling involves several stages and without dedicated classes for the associated data it would be difficult to keep track of the intermediate model inputs and outputs.

lulccR is substantially different from alternative environmental modelling frameworks. Most significantly, *lulccR* is designed for land use change modelling only, whereas frameworks such as PCRaster and TerraME provide general tools that can be applied to different spatial analysis problems such as land use change, hydrology and ecology. As a result, these tools are targeted towards the model developer rather than the end user. In contrast, existing land use change models are designed with the user in mind, with very few models providing any way for developers to improve or even understand the model implementation. With *lulccR* we have attempted to reduce the gap between user and developer. The R system is well suited for this task, as Pebesma et al. (2012) notes "the step from being a user to becoming a developer is small with R". The package system ensures that *lulccR* will work across Windows, Mac OS X and Unix platforms, whereas many existing applications are platform dependent. Comprehensive documentation of the functions, classes and methods of *lulccR*, together with complete working examples, enable the user to immediately start using the software, while the object-oriented design ensures that developers can easily write extensions to the package.

Despite its manifest advantages, there remain some drawbacks to land use change modelling in R. Firstly, the lack of a spatio-temporal database backend to support larger datasets (Gebbert and Pebesma, 2014) restricts the amount of data that can be used in a given application because R loads all data into memory. The *raster* package overcomes this limitation by storing raster files on disk and processing data in chunks (Hij-

mans, 2014). *lulccR* has been designed to make use of this facility where possible, however, during allocation it is necessary to load the values of several maps into the R workspace at once because the allocation procedure must consider every cell eligible for change simultaneously. The generic `predict` function belonging to the *raster* package provides one possible solution to this problem, allowing users to make predictions with predictive models in a memory-safe way. In effect, this would mean spatially explicit input data including observed land use maps and predictor variables could be handled in chunks and only the resulting probability surface would have to be loaded into the R workspace. However, this is not currently implemented in *lulccR* because it is excessively time consuming compared to the current approach. Despite this limitation, since most applications involve a relatively small geographic extent or, in the case of regional studies (e.g. Verburg and Overmars, 2009; Fuchs et al., 2015), use a coarser map resolution, memory should not normally cause *lulccR* applications to fail.

The software presented here is still in its infancy and there are several areas for improvement. The present allocation routines receive the quantity of land use change for each timestep before the allocation procedure begins. However, some recent models do not impose the quantity of change but instead allow change to occur stochastically based on land use suitability. For example, StocModLcc (Rosa et al., 2013) deforests a cell if the probability of deforestation is less than a random number from a uniform distribution. The quantity of change is simply the number of cells deforested after each cell in the study region is considered for deforestation twice, with the probability of change, which depends on the location of previous deforestation events, updated after the first round. One advantage of this approach is that it accounts for uncertainty in the quantity and location of change simultaneously, whereas the current routines in *lulccR* only consider the location of change as a stochastic process. Other models such as LandSHIFT (Schaldach et al., 2011) receive demand at the national or regional level from integrated assessment models such as IMAGE (Stehfast et al., 2014) or Nexus Land-Use (Souty et al., 2012). Coupling *lulccR* with this class of model would be a valuable addition to the software because land use change is increasingly recognised as a regional and global issue that occurs over multiple scales.

One of the main strengths of *lulccR* is that multiple model structures can be explored within the same environment. Thus, the more allocation routines available in the package the more useful it becomes. Two existing land use change models, StocModLCC and SIMLANDER, are written in R and available as open source software. Future work could integrate these routines with *lulccR* to broaden the different model structures and, therefore, improve the ability of *lulccR* to capture model structural uncertainty. The methods in the current version of *lulccR* only permit an inductive approach to land use change modelling. Deductive models are fundamentally different because they attempt to model explicitly the processes that drive land use change (Pérez-Vega et al., 2012). The main advantage of these models is that they are able to establish causality because they allow modellers to test specific theories about the location of change and predictor variables whereas inductive models simply associate land use change with explanatory variables through predictive models (Overmars et al., 2007). For example, the application for Plum Island Ecosystems shows that the presence of urban land is related to elevation, slope and distance to built land in 1971, however, the allocation models require no specific theory as to why this may be the case. Providing this class of model would permit multiscale studies whereby inductive and deductive land use change models operating at different spatial resolutions are dynamically coupled in order to better capture the complexity of the land use system (Moreira et al., 2009).

Free and open source software encourages the reproducibility of scientific results and allows users to adapt and extend code for their own purposes. Thus, we encourage the land use change community to participate in the future development of *lulccR*. Perhaps one of the simplest ways to improve the package is to experiment with the example datasets to identify bugs and areas for improvement. Those with more programming experience may wish to extend the functionality of the package themselves and contribute these changes upstream. In addition, existing land use change models can easily be included in the package by wrapping the original source code in R;

a straightforward task for commonly used compiled languages (C/C++, Fortran). Of course, users may also develop their own R packages that depend on *lulccR* for some functionality: this is one of the strengths of the R package system. Finally, we invite land use change modellers to submit land use change datasets (observed and, if possible, modelled land use maps and spatially explicit predictor variables) for inclusion in the package.

5 Conclusions

Land use change models are useful for several tasks, from supporting local planning decisions to studies of regional and global environmental change. However, currently available software for land use change modelling is generally closed-source and usually implements only one land use change model. In this paper we have presented *lulccR*, a free and open source software package providing an object-oriented framework for land use change modelling in R. *lulccR* allows the entire modelling process to be performed within the same environment, supports three different types of predictive model and includes two allocation routines. Releasing the software under an open source licence (GPL) means that users have access to the algorithms they implement when they run a particular model. As a result, they are able to identify improvements to the code and, under the terms of the licence, are free to redistribute these changes to the wider community. We view *lulccR* as an initial step towards an open paradigm for land use change modelling and hope, therefore, that the community will participate in its development.

Code availability

The *lulccR* source code currently resides on GitHub: https://github.com/simonmoulds/r_lulccR.

Discussion Paper | Discussion Paper | Discussion Paper | Discussion Paper |

The Supplement related to this article is available online at doi:10.5194/gmdd-8-3359-2015-supplement.

Acknowledgements. W. Buytaert and A. Mijic acknowledge support of the UK Natural Environment Research Council (contract NE/I022558/1). S. Moulds acknowledges support of the Grantham Institute for Climate Change, Imperial College London.

References

Alber, M., Reed, D., and McGlathery, K.: Coastal long term ecological research: introduction to the special issue, Oceanography, 26, 14–17, doi:10.5670/oceanog.2013.40, 2013. 3367

Aldwaik, S. Z. and Pontius, R. G.: Intensity analysis to unify measurements of size and stationarity of land changes by interval, category, and transition, Landscape Urban Plan., 106, 103–114, doi:10.1016/j.landurbplan.2012.02.010, 2012. 3367

Beale, C. M., Lennon, J. J., Yearsley, J. M., Brewer, M. J., and Elston, D. A.: Regression analysis of spatial data, Ecol. Lett., 13, 246–264, doi:10.1111/j.1461-0248.2009.01422.x, 2010. 3371

Bivand, R. S., Pebesma, E., and Gomez-Rubio, V.: Applied Spatial Data Analysis with R, 2nd edn., Springer, NY, available at: http://www.asdar-book.org/, 2013. 3364

Bivand, R. S., Keitt, T., and Rowlingson, B.: rgdal: Bindings for the Geospatial Data Abstraction Library, available at: http://CRAN.R-project.org/package=rgdal (last access: 16 April 2015), r package version 0.8-16, 2014. 3364

Boysen, L. R., Brovkin, V., Arora, V. K., Cadule, P., de Noblet-Ducoudré, N., Kato, E., Pongratz, J., and Gayler, V.: Global and regional effects of land-use change on climate in 21st century simulations with interactive carbon cycle, Earth Syst. Dynam., 5, 309–319, doi:10.5194/esd-5-309-2014, 2014. 3361

Brown, D. G., Verburg, P. H., Pontius, R. G., and Lange, M. D.: Opportunities to improve impact, integration, and evaluation of land change models, Current Opinion in Environmental Sustainability, 5, 452–457, doi:10.1016/j.cosust.2013.07.012, 2013. 3362

Cai, Y., Judd, K. L., and Lontzek, T. S.: Open science is necessary, Nature Climate Change, 2, 299–299, 2012. 3362

Câmara, G., Vinhas, L., Ferreira, K. R., De Queiroz, G. R., De Souza, R. C. M., Monteiro, A. M. V., De Carvalho, M. T., Casanova, M. A., and De Freitas, U. M.: TerraLib: an

open source GIS library for large-scale environmental and socio-economic applications, in: Open Source Approaches in Spatial Data Handling, 247–270, Springer, 2008. 3364

Carneiro, T. G. d. S., Andrade, P. R. d., Câmara, G., Monteiro, A. M. V., and Pereira, R. R.: An extensible toolbox for modeling nature–society interactions, Environ. Modell. Softw., 46, 104–117, doi:10.1016/j.envsoft.2013.03.002, 2013. 3363, 3364

Castella, J. and Verburg, P. H.: Combination of process-oriented and pattern-oriented models of land-use change in a mountain area of Vietnam, Ecol. Model., 202, 410–420, doi:10.1016/j.ecolmodel.2006.11.011, 2007. 3361, 3374, 3376

Chambers, J. M.: Programming with Data: a Guide to the S Language, Springer, 1998. 3366

Chambers, J. M.: Users, programmers, and statistical software, J. Comput. Graph. Stat., 9, 404–422, doi:10.1080/10618600.2000.10474890, 2000. 3366

Chambers, J. M.: Software for Data Analysis: Programming with R, Springer, 2008. 3364, 3366

Claes, M., Mens, T., and Grosjean, P.: On the maintainability of CRAN packages, in: 2014 Software Evolution Week – IEEE Conference on Software Maintenance, Reengineering and Reverse Engineering (CSMR-WCRE), Antwerp, 3–6 February, 308–312, IEEE, available at: http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=6747183 (last access: 16 April 2015), 2014. 3365

Couclelis, H.: "Where has the future gone?" Rethinking the role of integrated land-use models in spatial planning, Environ. Plann. A, 37, 1353–1371, doi:10.1068/a3785, 2005. 3361

Dormann, C. F., McPherson, J. M., Araújo, M. B., Bivand, R. S., Bolliger, J., Carl, G., Davies, R. G., Hirzel, A., Jetz, W., Kissling, W. D., Kühn, I., Ohlemüller, R., Peres-Neto, P. R., Reineking, B., Schröder, B., Schurr, F. M., and Wilson, R.: Methods to account for spatial autocorrelation in the analysis of species distributional data: a review, Ecography, 30, 609–628, 2007. 3371

Echeverria, C., Coomes, D. A., Hall, M., and Newton, A. C.: Spatially explicit models to analyze forest loss and fragmentation between 1976 and 2020 in southern Chile, Ecol. Model., 212, 439–449, doi:10.1016/j.ecolmodel.2007.10.045, 2008. 3371

Fiske, I. and Chandler, R.: unmarked: an R package for fitting hierarchical models of wildlife occurrence and abundance, J. Stat. Softw., 43, 1–23, 2011. 3364, 3365

Foley, J. A.: Global consequences of land use, Science, 309, 570–574, doi:10.1126/science.1111772, 2005. 3360

Fuchs, R., Herold, M., Verburg, P. H., and Clevers, J. G. P. W.: A high-resolution and harmonized model approach for reconstructing and analysing historic land changes in Europe, Biogeosciences, 10, 1543–1559, doi:10.5194/bg-10-1543-2013, 2013. 3373, 3374, 3378

Discussion Paper | Discussion Paper | Discussion Paper | Discussion Paper |

Fuchs, R., Herold, M., Verburg, P. H., Clevers, J. G., and Eberle, J.: Gross changes in reconstructions of historic land cover/use for Europe between 1900 and 2010, Glob. Change Biol., 21, 299–313, doi:10.1111/gcb.12714, 2015. 3382

Gebbert, S. and Pebesma, E.: A temporal GIS for field based environmental modeling, Environ. Modell. Softw., 53, 1–12, doi:10.1016/j.envsoft.2013.11.001, 2014. 3381

Grenouillet, G., Buisson, L., Casajus, N., and Lek, S.: Ensemble modelling of species distribution: the effects of geographical and environmental ranges, Ecography, 34, 9–17, doi:10.1111/j.1600-0587.2010.06152.x, 2011. 3363

Hewitt, R., Díaz Pacheco, J., and Moya Gómez, B.: A cellular automata land use model for the R software environment, available at: http://simlander.wordpress.com/ (last access: 11 January 2015), 2013. 3364

Hijmans, R. J.: raster: Geographic data analysis and modeling, available at: http://CRAN.R-project.org/package=raster (last access: 16 April 2015), r package version 2.2-31, 2014. 3364

Hobbie, J. E., Carpenter, S. R., Grimm, N. B., Gosz, J. R., and Seastedt, T. R.: The US long term ecological research program, BioScience, 53, 21–32, 2003. 3367

Ince, D. C., Hatton, L., and Graham-Cumming, J.: The case for open computer programs, Nature, 482, 485–488, doi:10.1038/nature10836, 2012. 3362

Joppa, L. N., McInerny, G., Harper, R., Salido, L., Takeda, K., O'Hara, K., Gavaghan, D., and Emmott, S.: Troubling trends in scientific software use, Science, 340, 814–815, 2013. 3366

Knutti, R. and Sedláček, J.: Robustness and uncertainties in the new CMIP5 climate model projections, Nature Climate Change, 3, 369–373, doi:10.1038/nclimate1716, 2012. 3363

Kuhn, M., Wing, J., Weston, S., Williams, A., Keefer, C., and Engelhardt, A.: caret: Classification and regression training, available at: http://CRAN.R-project.org/package=caret (last access: 16 April 2015), r package version 5.15-044, 2012. 3371

Le Quéré, C., Raupach, M. R., Canadell, J. G., Marland, G., Raupach, M. R., Canadell, J. G., Marland, G., Bopp, L., Ciais, P., Conway, T. J., Doney, S. C., Feely, R. A., Foster, P., Friedlingstein, P., Gurney, K., Houghton, R. A., House, J. I., Huntingford, C., Levy, P. E., Lomas, M. R., Majkut, J., Metzl, N., Ometto, J. P., Peters, G. P., Prentice, I. C., Randerson, J. T., Running, S. W., Sarmiento, J. L., Schuster, U., Sitch, S., Takahashi, T., Viovy, N., van der Werf, G. R., and Woodward, F. I.: Trends in the sources and sinks of carbon dioxide, Nat. Geosci., 2, 831–836, doi:10.1038/ngeo689, 2009. 3361

Li, K., Coe, M., Ramankutty, N., and Jong, R. D.: Modeling the hydrological impact of land-use change in West Africa, J. Hydrol., 337, 258–268, doi:10.1016/j.jhydrol.2007.01.038, 2007. 3361

Liaw, A. and Wiener, M.: Classification and Regression by randomForest, R news, 2, 18–22, available at: ftp://131.252.97.79/Transfer/Treg/WFRE_Articles/Liaw_02_Classification%20and%20regression%20by%20randomForest.pdf (last access: 16 April 2015), 2002. 3371

Lin, Y., Wu, P., and Hong, N.: The effects of changing the resolution of land-use modeling on simulations of land-use patterns and hydrology for a watershed land-use planning assessment in Wu-Tu, Taiwan, Landscape Urban Plan., 87, 54–66, doi:10.1016/j.landurbplan.2008.04.006, 2008. 3361

Magliocca, N. R. and Ellis, E. C.: Using Pattern-oriented Modeling (POM) to Cope with uncertainty in multi-scale agent-based models of land change: POM in Multi-scale ABMs of land change, Transactions in GIS, 17, 883–900, doi:10.1111/tgis.12012, 2013. 3361

Mas, J., Kolb, M., Paegelow, M., Camacho Olmedo, M. T., and Houet, T.: Inductive pattern-based land use/cover change models: a comparison of four software packages, Environ. Modell. Softw., 51, 94–111, doi:10.1016/j.envsoft.2013.09.010, 2014. 3361, 3362, 3363, 3374, 3395

Mascaro, J., Asner, G. P., Knapp, D. E., Kennedy-Bowdoin, T., Martin, R. E., Anderson, C., Higgins, M., and Chadwick, K. D.: A tale of two "Forests": random forest machine learning aids tropical forest carbon mapping, PLoS ONE, 9, e85993, doi:10.1371/journal.pone.0085993,2014. 3371

MassGIS: Massachusetts Geographic Information System, MassGIS, available at: http://www.mass.gov/anf/research-and-tech/it-serv-and-support/application-serv/office-of-geographic-information-massgis/ (last access: 16 April 2015), 2015. 3367

Moreira, E., Costa, S., Aguiar, A. P., Câmara, G., and Carneiro, T.: Dynamical coupling of multi-scale land change models, Landscape Ecol., 24, 1183–1194, doi:10.1007/s10980-009-9397-x, 2009. 3361, 3364, 3383

Morin, A., Urban, J., Adams, P. D., Foster, I., Sali, A., Baker, D., and Sliz, P.: Shining light into black boxes, Science, 336, 159–160, 2012. 3362

Morse, N. B. and Wollheim, W. M.: Climate variability masks the impacts of land use change on nutrient export in a suburbanizing watershed, Biogeochemistry, 121, 45–59, doi:10.1007/s10533-014-9998-6, 2014. 3367

Mulia, R., Widayati, A., Putra Agung, S., and Zulkarnain, M. T.: Low carbon emission development strategies for Jambi, Indonesia: simulation and trade-off analysis using the FALLOW model, Mitigation and Adaptation Strategies for Global Change, 19, 773–788, doi:10.1007/s11027-013-9485-8, 2014. 3363

Nelson, E., Sander, H., Hawthorne, P., Conte, M., Ennaanay, D., Wolny, S., Manson, S., and Polasky, S.: Projecting global land-use change and its effect on ecosystem service provision and biodiversity with simple models, PLoS ONE, 5, e14327, doi:10.1371/journal.pone.0014327, 2010. 3361

Overmars, K., de Koning, G., and Veldkamp, A.: Spatial autocorrelation in multi-scale land use models, Ecol. Model., 164, 257–270, doi:10.1016/S0304-3800(03)00070-X, 2003. 3371

Overmars, K. P., Verburg, P. H., and Veldkamp, A.: Comparison of a deductive and an inductive approach to specify land suitability in a spatially explicit land use model, Land Use Policy, 24, 584–599, doi:10.1016/j.landusepol.2005.09.008, 2007. 3361, 3383

Pebesma, E. J. and Bivand, R. S.: Classes and methods for spatial data in R, R News, 5, 9–13, available at: http://CRAN.R-project.org/doc/Rnews/ (last access: 16 April 2015), 2005. 3364

Pebesma, E. J., Nüst, D., and Bivand, R.: The R software environment in reproducible geoscientific research, EOS T. Am. Geophys. Un., 93, 163–163, 2012. 3362, 3365, 3381

Peng, R. D.: Reproducible research in computational science, Science, 334, 1226–1227, doi:10.1126/science.1213847, 2011. 3362

Pérez-Vega, A., Mas, J., and Ligmann-Zielinska, A.: Comparing two approaches to land use/cover change modeling and their implications for the assessment of biodiversity loss in a deciduous tropical forest, Environ. Modell. Softw., 29, 11–23, doi:10.1016/j.envsoft.2011.09.011, 2012. 3363, 3383

Petzoldt, T. and Rinke, K.: Simecol: an object-oriented framework for ecological modeling in R, J. Stat. Softw., 22, 1–31, 2007. 3364

Pitman, A. J., de Noblet-Ducoudré, N., Cruz, F. T., Davin, E. L., Bonan, G. B., Brovkin, V., Claussen, M., Delire, C., Ganzeveld, L., Gayler, V., van den Hurk, B. J. J. M., Lawrence, P. J., van der Molen, M. K., Müller, C., Reick, C. H., Seneviratne, S. I., Strengers, B. J., and Voldoire, A.: Uncertainties in climate responses to past land cover change: first results from the LUCID intercomparison study, Geophys. Res. Lett., 36, L14814, doi:10.1029/2009GL039076, 2009. 3361

Pontius, R. G. and Parmentier, B.: Recommendations for using the relative operating characteristic (ROC), Landscape Ecol., 367–382, doi:10.1007/s10980-013-9984-8, 2014. 3367, 3372

Pontius, R. G. and Schneider, L. C.: Land-cover change model validation by an ROC method for the Ipswich watershed, Massachusetts, USA, Agr. Ecosyst. Environ., 85, 239–248, 2001. 3370, 3374

Pontius, R. G and Spencer, J.: Uncertainty in extrapolations of predictive land-change models, Environ. Plann. B, 32, 211–230, doi:10.1068/b31152, 2005. 3363

Pontius, R. G., Cornell, J. D., and Hall, C. A.: Modeling the spatial pattern of land-use change with GEOMOD2: application and validation for Costa Rica, Agr. Ecosyst. Environ., 85, 191–203, 2001.

Pontius, R. G., Huffaker, D., and Denman, K.: Useful techniques of validation for spatially explicit land-change models, Ecol. Model., 179, 445–461, doi:10.1016/j.ecolmodel.2004.05.010, 2004a. 3369

Pontius, R. G., Shusas, E., and McEachern, M.: Detecting important categorical land changes while accounting for persistence, Agr. Ecosyst. Environ., 101, 251–268, doi:10.1016/j.agee.2003.09.008, 2004b. 3369, 3379

Pontius, R. G., Boersma, W., Castella, J., Clarke, K., Nijs, T., Dietzel, C., Duan, Z., Fotsing, E., Goldstein, N., Kok, K., Koomen, E., Lippitt, C. D., McConnell, W., Mohd Sood, A., Pijanowski, B., Pithadia, S., Sweeney, S., Trung, T. N., Veldkamp, A. T., and Verburg, P. H.: Comparing the input, output, and validation maps for several models of land change, Ann. Regional Sci., 42, 11–37, doi:10.1007/s00168-007-0138-2, 2007. 3379

Pontius, R. G., Peethambaram, S., and Castella, J.: Comparison of three maps at multiple resolutions: a case study of land change simulation in Cho Don district, Vietnam, Ann. Assoc. Am. Geogr., 101, 45–62, doi:10.1080/00045608.2010.517742, 2011. 3379, 3380

Prudhomme, C., Giuntoli, I., Robinson, E. L., Clark, D. B., Arnell, N. W., Dankers, R., Fekete, B. M., Franssen, W., Gerten, D., Gosling, S. N., Hagemann, S., Hannah, D. M., Kim, H., Masaki, Y., Satoh, Y., Stacke, T., Wada, Y., and Wisser, D.: Hydrological droughts in the 21st century, hotspots and uncertainties from a global multimodel ensemble experiment, P. Natl. Acad. Sci. USA, 111, 3262–3267, doi:10.1073/pnas.1222473110, 2014. 3363

R Core Team: R: A Language and Environment for Statistical Computing, R Foundation for Statistical Computing, Vienna, Austria, available at: http://www.R-project.org/ (last access: 16 April 2015), 2014. 3364

Ray, D. K. and Pijanowski, B. C.: A backcast land use change model to generate past land use maps: application and validation at the Muskegon River watershed of Michigan, USA, Journal of Land Use Science, 5, 1–29, doi:10.1080/17474230903150799, 2010. 3374

Rodell, M., Velicogna, I., and Famiglietti, J. S.: Satellite-based estimates of groundwater depletion in India, Nature, 460, 999–1002, doi:10.1038/nature08238, 2009. 3361

Rodríguez Eraso, N., Armenteras-Pascual, D., and Alumbreros, J. R.: Land use and land cover change in the Colombian Andes: dynamics and future scenarios, Journal of Land Use Science, 8, 154–174, doi:10.1080/1747423X.2011.650228, 2013. 3361

Rosa, I. M. D., Purves, D., Souza, C., and Ewers, R. M.: Predictive modelling of contagious deforestation in the Brazilian Amazon, PLoS ONE, 8, e77231, doi:10.1371/journal.pone.0077231, 2013. 3361, 3364, 3373, 3382

Rosa, I. M. D., Ahmed, S. E., and Ewers, R. M.: The transparency, reliability and utility of tropical rainforest land-use and land-cover change models, Glob. Change Biol., 20, 1707–1722, doi:10.1111/gcb.12523, 2014. 3362, 3363, 3367

Schaldach, R., Alcamo, J., Koch, J., Kölking, C., Lapola, D. M., Schüngel, J., and Priess, J. A.: An integrated approach to modelling land-use change on continental and global scales, Environ. Modell. Softw., 26, 1041–1051, doi:10.1016/j.envsoft.2011.02.013, 2011. 3362, 3382

Schmitz, O., Karssenberg, D., van Deursen, W., and Wesseling, C.: Linking external components to a spatio-temporal modelling framework: coupling MODFLOW and PCRaster, Environ. Modell. Softw., 24, 1088–1099, doi:10.1016/j.envsoft.2009.02.018, available at: http://linkinghub.elsevier.com/retrieve/pii/S1364815209000516, 2009. 3363

Seneviratne, S. I., Corti, T., Davin, E. L., Hirschi, M., Jaeger, E. B., Lehner, I., Orlowsky, B., and Teuling, A. J.: Investigating soil moisture–climate interactions in a changing climate: a review, Earth-Sci. Rev., 99, 125–161, doi:10.1016/j.earscirev.2010.02.004, 2010. 3361

Shankar, P. V., Kulkarni, H., and Krishnan, S.: India's groundwater challenge and the way forward, Econ. Polit. Weekly, 46, 37–45, 2011. 3361

Sing, T., Sander, O., Beerenwinkel, N., and Lengauer, T.: ROCR: visualizing classifier performance in R, Bioinformatics, 21, 3940–3941, doi:10.1093/bioinformatics/bti623, 2005. 3372

Soares-Filho, B. S., Coutinho Cerqueira, G., and Lopes Pennachin, C.: DINAMICA-a stochastic cellular automata model designed to simulate the landscape dynamics in an Amazonian colonization frontier, Ecol. Model., 154, 217–235, 2002. 3362

Sohl, T. L., Sayler, K. L., Drummond, M. A., and Loveland, T. R.: The FORE-SCE model: a practical approach for projecting land cover change using scenario-based modeling, Journal of Land Use Science, 2, 103–126, doi:10.1080/17474230701218202, 2007. 3361, 3374

Sohl, T. L., Sleeter, B. M., Sayler, K. L., Bouchard, M. A., Reker, R. R., Bennett, S. L., Sleeter, R. R., Kanengieter, R. L., and Zhu, Z.: Spatially explicit land-use and land-cover scenarios for the Great Plains of the United States, Agr. Ecosyst. Environ., 153, 1–15, doi:10.1016/j.agee.2012.02.019, 2012. 3361

Souty, F., Brunelle, T., Dumas, P., Dorin, B., Ciais, P., Crassous, R., Müller, C., and Bondeau, A.: The Nexus Land-Use model version 1.0, an approach articulating biophysical potentials and economic dynamics to model competition for land-use, Geosci. Model Dev., 5, 1297–1322, doi:10.5194/gmd-5-1297-2012, 2012. 3361, 3382

Stehfast, E., van Vuuren, D., Kram, T., Bouwman, L., Alkemade, R., Bakkenes, M., Biemans, H., Bouwman, A., den Elzen, M., Janse, J., Lucas, P., van Minnen, J., Muller, M., and Prins, A. G.: Integrated Assessment of Global Environmental Change with IMAGE 3.0 – Model Description and Policy Applications, available at: http://www.pbl.nl/en/publications/integrated-assessment-of-global-environmental-change-with-IMAGE-3.0 (last access: 16 April 2015), iSBN 978-94-91506-71-0, 2014. 3382

Steiniger, S. and Hunter, A. J.: The 2012 free and open source GIS software map – a guide to facilitate research, development, and adoption, Comput. Environ. Urban, 39, 136–150, doi:10.1016/j.compenvurbsys.2012.10.003, 2013. 3362

Tallaksen, L. M. and Stahl, K.: Spatial and temporal patterns of large-scale droughts in Europe: model dispersion and performance: Tallaksen and Stahl: large-scale hydrological droughts, Geophys. Res. Lett., 41, 429–434, doi:10.1002/2013GL058573, 2014. 3363

Taylor, K. E., Stouffer, R. J., and Meehl, G. A.: An overview of CMIP5 and the experiment design, B. Am. Meteorol. Soc., 93, 485–498, doi:10.1175/BAMS-D-11-00094.1, 2012. 3363

Tayyebi, A., Pijanowski, B. C., Linderman, M., and Gratton, C.: Comparing three global parametric and local non-parametric models to simulate land use change in diverse areas of the world, Environ. Modell. Softw., 59, 202–221, doi:10.1016/j.envsoft.2014.05.022, 2014. 3370

Therneau, T., Atkinson, B., and Ripley, B.: rpart: Recursive Partitioning and Regression Trees, available at: http://CRAN.R-project.org/package=rpart (last access: 16 April 2015), r package version 4.1-8, 2014. 3370

Discussion Paper | Discussion Paper | Discussion Paper | Discussion Paper |

Turner, B. L., Lambin, E. F., and Reenberg, A.: The emergence of land change science for global environmental change and sustainability, P. Natl. Acad. Sci. USA, 104, 20666–20671, 2007. 3360

van Noordwijk, M.: Scaling trade-offs between crop productivity, carbon stocks and biodiversity in shifting cultivation landscape mosaics: the FALLOW model, Ecol. Model., 149, 113–126, 2002. 3363

van Vliet, J., Bregt, A. K., and Hagen-Zanker, A.: Revisiting Kappa to account for change in the accuracy assessment of land-use change models, Ecol. Model., 222, 1367–1375, doi:10.1016/j.ecolmodel.2011.01.017, 2011. 3379

Veldkamp, A. and Fresco, L.: CLUE: a conceptual model to study the conversion of land use and its effects, Ecol. Model., 85, 253–270, 1996. 3364

Veldkamp, A. and Lambin, E. F.: Predicting land-use change, Agr. Ecosyst. Environ., 85, 1–6, 2001. 3361

Verburg, P. H. and Overmars, K. P.: Combining top-down and bottom-up dynamics in land use modeling: exploring the future of abandoned farmlands in Europe with the Dyna-CLUE model, Landscape Ecol., 24, 1167–1181, doi:10.1007/s10980-009-9355-7, 2009. 3361, 3362, 3375, 3382

Verburg, P. H., De Koning, G. H. J., Kok, K., Veldkamp, A., and Bouma, J.: A spatial explicit allocation procedure for modelling the pattern of land use change based upon actual land use, Ecol. Model., 116, 45–61, 1999. 3364

Verburg, P. H., Soepboer, W., Veldkamp, A., Limpiada, R., Espaldon, V., and Mastura, S. S.: Modeling the spatial dynamics of regional land use: the CLUE-S model, Environ. Manage., 30, 391–405, doi:10.1007/s00267-002-2630-x, 2002. 3361, 3362, 3368, 3370, 3371, 3374, 3375, 3376

Verburg, P. H., Eck, J. R. R. v., Nijs, T. C. M. D., Dijst, M. J., and Schot, P.: Determinants of land-use change patterns in the Netherlands, Environ. Plann. B, 31, 125–150, doi:10.1068/b307, 2004. 3362, 3368

Verburg, P. H., Tabeau, A., and Hatna, E.: Assessing spatial uncertainties of land allocation using a scenario approach and sensitivity analysis: a study for land use in Europe, J. Environ. Manage., 127, S132–S144, doi:10.1016/j.jenvman.2012.08.038, 2013. 3363

Villamor, G. B. and Lasco, R. D.: Rewarding upland people for forest conservation: experience and lessons learned from case studies in the Philippines, Journal of Sustainable Forestry, 28, 304–321, doi:10.1080/10549810902791499, 2009. 3368

Wada, Y., van Beek, L. P. H., and Bierkens, M. F. P.: Nonsustainable groundwater sustaining irrigation: a global assessment, Water Resour. Res., 48, W00L06, doi:10.1029/2011WR010562, 2012. 3361

Wassenaar, T., Gerber, P., Verburg, P., Rosales, M., Ibrahim, M., and Steinfeld, H.: Projecting land use changes in the Neotropics: the geography of pasture expansion into forest, Global Environ. Chang., 17, 86–104, doi:10.1016/j.gloenvcha.2006.03.007, 2007. 3371

Wilson, G., Aruliah, D. A., Brown, C. T., Chue Hong, N. P., Davis, M., Guy, R. T., Haddock, S. H. D., Huff, K. D., Mitchell, I. M., Plumbley, M. D., Waugh, B., White, E. P., and Wilson, P.: Best practices for scientific computing, PLoS Biology, 12, e1001745, doi:10.1371/journal.pbio.1001745, 2014. 3366

Wu, D., Liu, J., Zhang, G., Ding, W., Wang, W., and Wang, R.: Incorporating spatial autocorrelation into cellular automata model: an application to the dynamics of Chinese tamarisk (*Tamarix chinensis* Lour.), Ecol. Model., 220, 3490–3498, doi:10.1016/j.ecolmodel.2009.03.008, 2009. 3371

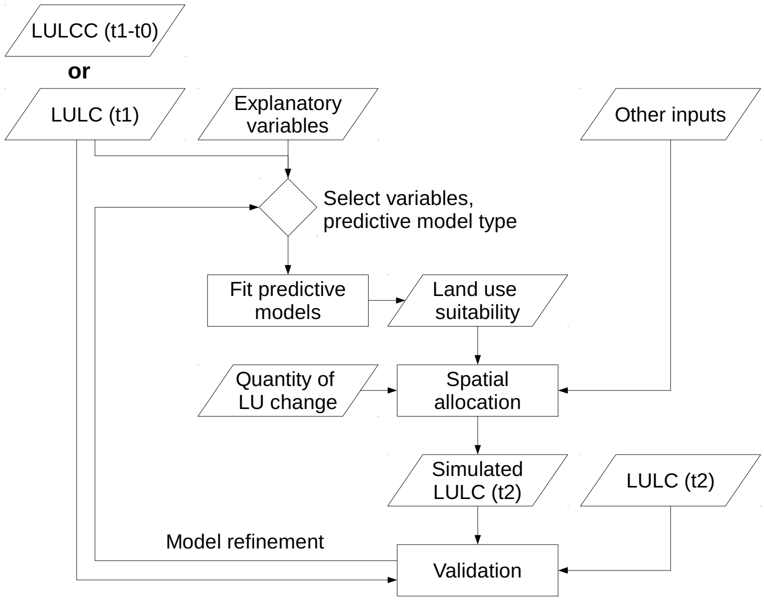

Figure 1. Diagram showing the general methodology used for inductive land use change modelling applications, adapted from Mas et al. (2014).

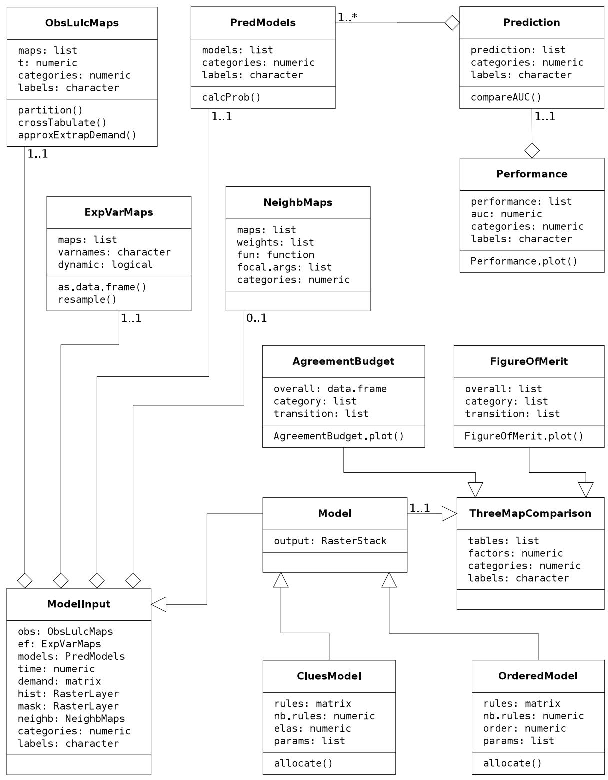

Figure 2. Class diagram in the Unified Modeling Language (UML) for *lulccR*, showing the main classes and core functions included in the package.

Discussion Paper | Discussion Paper | Discussion Paper | Discussion Paper |

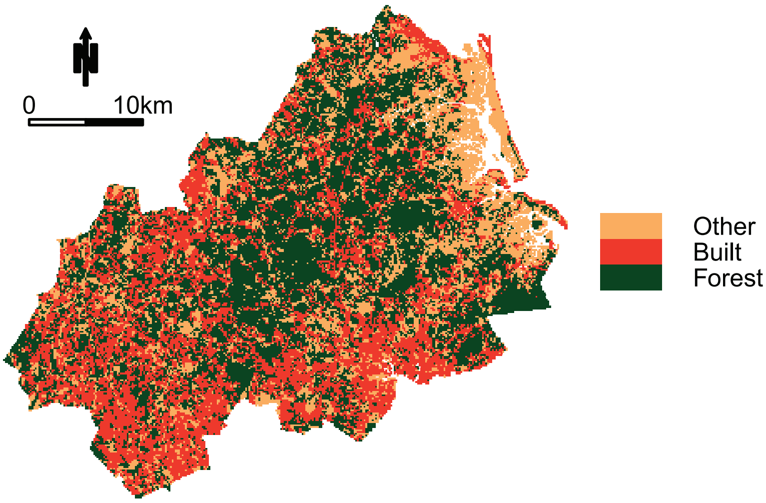

Figure 3. Plum Island Ecosystems site land use map for 1985. In recent years the site has undergone extensive land use change from forest to built areas.

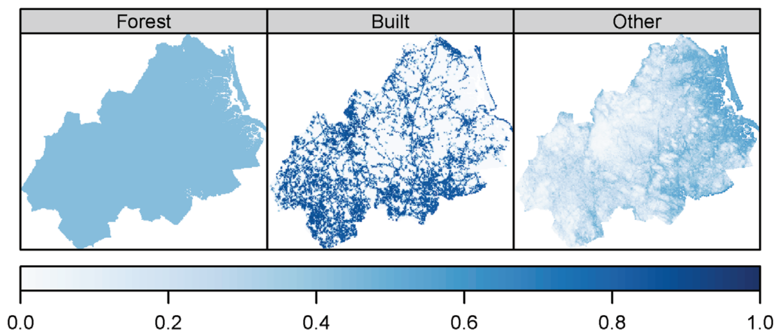

Figure 4. Land use suitability maps for Plum Islands Ecosystems study area. The "forest" land use class has uniform suitability because we employ a null model. Occurrence of "built" is related to elevation, slope and distance to 1971 built area, while "other" is related to elevation and slope only.

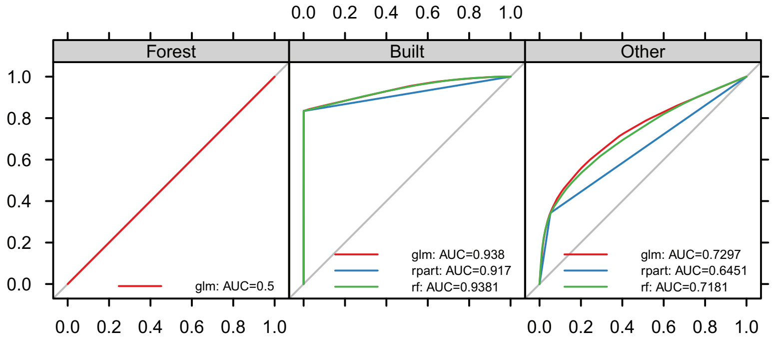

Figure 5. ROC curves for each statistical model for each land use. Note that because the forest class employs a null model only the logistic regression model is calculated.

Discussion Paper | Discussion Paper | Discussion Paper | Discussion Paper |

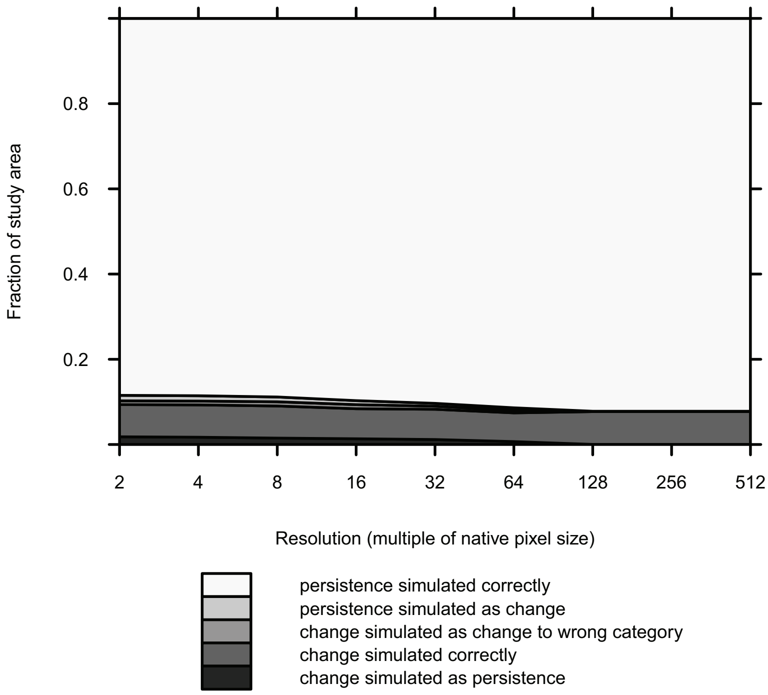

Figure 6. Overall agreement budget comparing the *lulccR* CLUE-S algorithm with the original model output for 2011. This shows a good level of agreement between the two maps: the proportion of persistence and change simulated correctly is high compared to incorrectly simulated persistence or change.

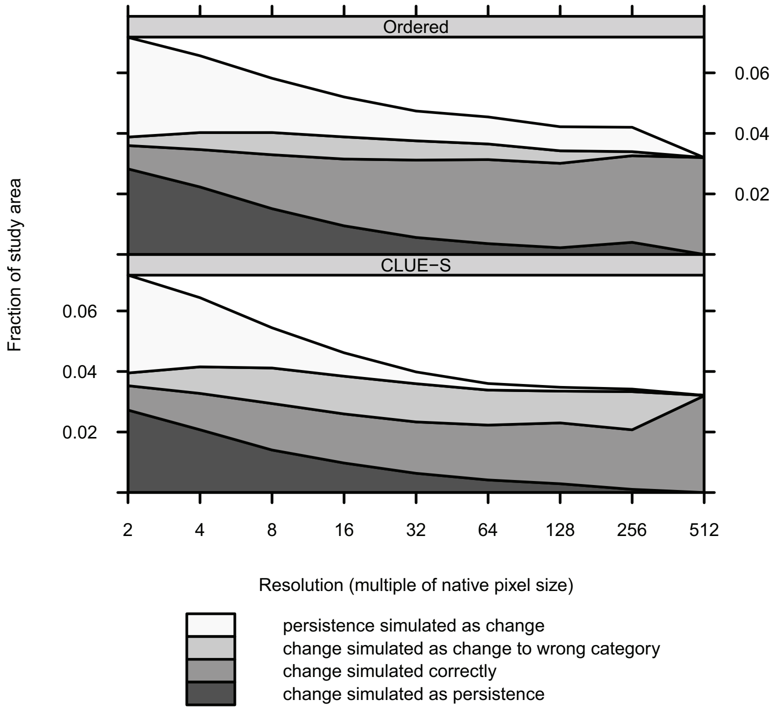

Figure 7. Agreement budget for the transition from “Forest” to “Built” for the two model outputs considering reference maps at 1985 and 1999 and simulated map for 1999. The plot shows the amount of correctly allocated change increases as the map resolution decreases.

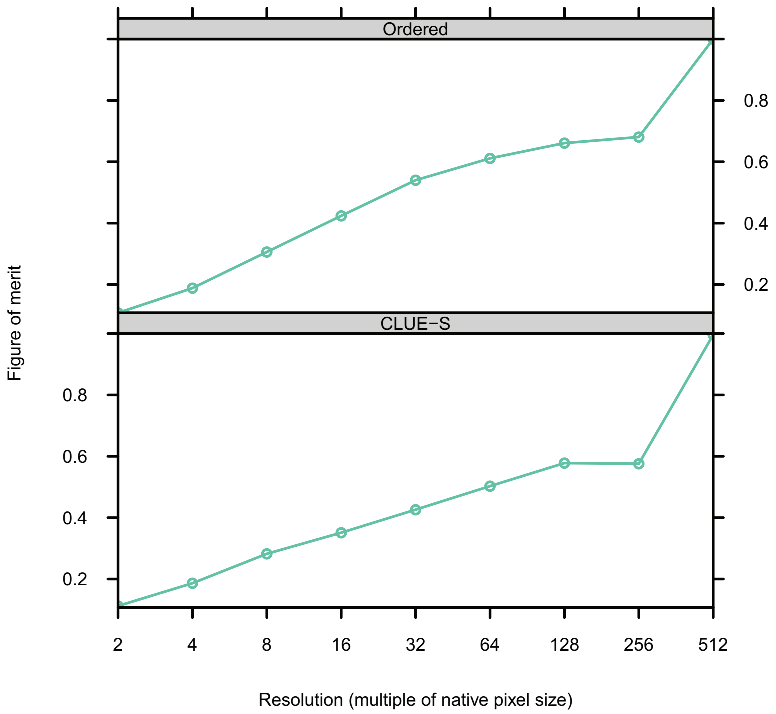

Figure 8. Figure of merit scores corresponding to the agreement budgets depicted in Fig. 7.