An open and extensible framework for spatially explicit land use change modelling: the lulcc R package

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

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

Abstract

We present the lulcc software 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
5 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 alternative models; (3) Additional
10 software is required because existing applications frequently perform only the spatial allocation of change. The package includes a stochastic ordered allocation procedure as well as an implementation of the 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 lulcc will enable future model development and
15 comparison within an open environment.

1 Introduction

Spatially explicit land use change models are used to understand and quantify key processes that affect land use and land cover change and simulate past and future change
20 (Veldkamp and Lambin, 2001; Mas et al., 2014). These models are commonly implemented 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 land use change modelling software to be made available (e.g. Verburg et al., 2002; Soares-Filho et al.,
25 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

2

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
[5] 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). In this paper we describe the development of **lulcc**, a new R package designed to foster an open approach to land use change science.

[10]

Current software packages for land use change modelling 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
[15] 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
[20] 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 modelling applications (Verburg et al., 2013). Under these circumstances it is useful to experiment with various models to identify the model that performs best in terms of
[25] 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). 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

3

prevent the widespread use of ensemble land use change predictions (Rosa et al., 2014). As a result, the uncertainty associated with model outcomes is rarely communicated in a formal way, raising questions about the utility of such models (Pontius and Spencer, 2005).

5   An alternative approach is to develop frameworks that allow several modelling approaches to be implemented within the same environment. One such application is PCRaster, 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
10   such as map algebra and neighbourhood functions. Alternatively, the PCRaster application programming interface (API) allows users to extend its functionality in various 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 land use change model, is built using the PCRaster framework. TerraME (Carneiro et al.,
15   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 various 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 several spatio-temporal data types, includes an
20   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 implementations of CLUE-S and its predecessor, CLUE (Veldkamp and Fresco, 1996; Verburg et al., 1999), written in Lua.

25   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

4

community, largely due to the **sp** package (Pebesma and Bivand, 2005; Bivand et al., 2013) which unified many alternative approaches for dealing with spatial data in R and allowed subsequent package developers to use a common framework for spatial analysis. The **raster** package (Hijmans, 2014) provides many functions for raster data manipulation

5 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.,

10 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. The remainder of this paper is divided into four sections. First, we discuss the principle design goals of **lulcc**. We then describe the software and demonstrate its main functionality

15 with an example application to the Plum Island Ecosystems site, using data included with the package. This is followed by a discussion of the strengths and main limitations of the software and approach, as well as areas for future development. Finally we draw brief conclusions from the project.

20 **2 Design goals**

The first design goal of **lulcc** is to provide a framework that allows users to perform various stages of the modelling process illustrated by Figure 1 within the same environment. It therefore includes methods to process and explore model input, fit and evaluate predictive models, allocate land use change spatially, validate the model and visualise model outputs.

25 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 seperate aspects of the procedure can easily be revisited. Thirdly, it is straightforward to experiment with different model setups. Finally, and perhaps most importantly, it improves the reproducibility of scientific results because the entire modelling process can be expressed programmatically and be communicated as such with reasonable effort (Pebesma et al., 2012).

lulcc is intended as an alternative to current paradigm of closed source, specialised software programs 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 datasets 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 statistical modelling, data management, spatial analysis and visualisation.

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 well documented software that is easy to use and accessible for a users with varying 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 **lulcc** is accessible so that users can locate the methods in use and understand algorithm implementations. Acknowledging that

6

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 various 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 several 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 **lulcc** 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 structure of **lulcc**, while Table 1 shows the functions included with the package. Here we describe the main components of **lulcc** integrated with an example application for the Plum Island Ecosystems dataset. The script used in this paper, including the code used to create the various figures, is supplied with the package as a "demo". Instructions to obtain the package and run the demo script are provided in the Code availability section.

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 **lulcc** package includes two datasets that have been widely used in the land use change

7

community, allowing users to quickly start exploring the modelling framework. The first of these contains data from the Plum Island Ecosystems Long Term Ecological Research site in northeast Massachusetts (http://pie-lter.ecosystems.mbl.edu/), which in recent decades has undergone extensive land use change from forest to residential use (Aldwaik and Pon-

5 tius, 2012). The dataset included in **lulcc** 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 together with maps of three predictor variables: elevation, slope and distance to built land in 1985. The second dataset includes information from Sibuyan Island in the Phillipines, and

10 is a modified version of the dataset supplied with the CLUE-S model (Verburg et al., 2002).

3.2 Data processing

One of the most challenging aspects of land use change modelling is to obtain and process the correct input data. Currently **lulcc** requires all spatially explicit input data to exist either in the file system, in any of the formats supported by **raster**, or in the R workspace as

15 **raster** objects (RasterLayer, RasterStack or RasterBrick). The most fundamental input required by land use change models is an initial map of observed land use, which is usually obtained from classified remotely sensed data. This map represents the initial condition for model simulations and, for inductive modelling, is used to fit predictive models. Sometimes it is more useful to consider observed land use transitions: in this case an additional map for

20 an earlier time point is required, as shown by Figure 1. Ideally, two more observed land use maps for subsequent time points should be obtained for calibrating and validating the land use change model (Pontius et al., 2004a). The current version of the software only supports categorical land use data, which means that each pixel must belong to exactly one category.

25 In **lulcc** observed land use data are represented by the ObsLulcRasterStack class. In the following code snippet we load the package into the current session, create an ObsLulcRasterStack object for the Plum Island Ecosystems dataset and plot the result (Figure 3):

8

```
> library(lulcc)
> data(pie)
> obs <- ObsLulcRasterStack(x=pie,
                            pattern="lu",
                            categories=c(1,2,3),
                            labels=c("Forest","Built","Other"),
                            t=c(0,6,14))
> plot(obs)
```

The ObsLulcRasterStack object is important to land use change studies in **lulcc** because it defines the spatial domain of subsequent operations. The t argument in the constructor function specifies the time points associated with the observed land use maps. The first time point must always be zero; if additional maps are present they should be associated with time points greater than zero, even in backcast models. In most land use change modelling applications the timestep between two time points 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 observed land use maps from two time points to identify the main historical transitions in the study region (Pontius et al., 2004b), which can be used as the basis for further research into the processes driving change. In **lulcc** we use the crossTabulate function for this purpose:

```
> crossTabulate(x=obs, times=c(0,14))
       Forest Built Other
Forest 44107  4250   656
Built     11 36957   154
Other   1259  2248 23921
```

9

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 allocation 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 **lulcc** these two types of explanatory variable are separated by a simple naming convention, which is explained in detail in the package documentation (see Supplementary material). Collectively, they are represented by an object of class ExpVarRasterList, which can be created as follows:

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

Apart from observed land use and explanatory variables other input maps may be required. The two allocation routines currently included with **lulcc** 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 **lulcc** as standard RasterLayer objects. All input maps should have the same spatial resolution as the corresponding ObsLulcRasterStack object. This can be achieved using the resample function from the **raster** package, which has been extended to receive lulcc objects. The ExpVarRasterList object created above can be resampled to the parameters of an ObsLulcRasterStack object with the following command:

```
> ef <- resample(ef, obs)
```

3.3 Predictive modelling

Inductive land use change models relate the pattern of observed land use to spatially explicit explanatory variables. Logistic regression is a common type of predictive model used for inductive land use change modelling (e.g. Pontius and Schneider, 2001; Verburg et al., 2002). However, there is growing interest in the application of local and non-parametric models (e.g. Tayyebi et al., 2014). One reason why R is attractive for land use change

10

modelling is that it has become the *de facto* standard for statistical software development. As a result, **lulcc** can easily support various predictive modelling techniques by utilising code from existing R packages. Currently, **lulcc** 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).

Parametric models such as logistic regression assume the data to be independent and identically distributed (Overmars et al., 2003). 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). There is also some evidence that non-parametric models may be affected by spatial autocorrelation Mascaro et al. (2014), even though they do not assume independence. A simple approach to reduce the impact of this phenomenon is to fit predictive models to a random subset of the data (e.g. Verburg et al., 2002; Wassenaar et al., 2007; Echeverria et al., 2008). In the following code snippet we create training and testing partitions for the Plum Island Ecosystems dataset by performing a stratified random sample. We do this using the map for 1985 to illustrate the procedure when only one observed map is available. We then extract the data for the training partition with the getPredictiveModelInputData function and pass the resulting data.frame to the three model fitting functions:

```
> part <- partition(x=obs[[1]], size=0.1, spatial=TRUE)
> train.data
        <- getPredictiveModelInputData(obs=obs,
                                        ef=ef,
                                        cells=part[["train"]],
                                        t=0)

> forms <- list(Built~ef_001+ef_002+ef_003,
```

11

```
                       Forest~ef_001+ef_002,
                       Other~ef_001+ef_002)

   > glm.models     <- glmModels(formula=forms,
5                                  family=binomial,
                                   data=train.data,
                                   obs=obs)

   > rpart.models  <- rpartModels(formula=forms,
10                                  data=train.data,
                                   obs=obs)

   > rf.models      <- randomForestModels(formula=forms,
                                           data=train.data,
15                                          obs=obs)
```

The model fitting functions each return an object of class PredictiveModelList containing a predictive model for each land use type. With these objects it straightforward to map the suitability of every pixel in the study region to the various land uses. To do this, we use the generic predict function with some additional functionality from the **raster** package and plot the resulting RasterStack object (Figure 4):

```
   > all.data <- as.data.frame(x=ef, cells=part[["all"]])
   > probmaps <- predict(object=glm.models,
                          newdata=all.data,
                          data.frame=TRUE)
25 > points    <- rasterToPoints(obs[[1]], spatial=TRUE)
   > probmaps <- SpatialPointsDataFrame(points, probmaps)
   > probmaps <- rasterize(x=probmaps, y=obs[[1]],
                           field=names(probmaps))
   > levelplot(probmaps)
```

12

In some circumstances it may be appropriate to supply a model with no explanatory variables to an allocation routine. For example, Verburg and Overmars (2009) used such a model for natural and semi-natural vegetation because in their particular case study the selection of pixels for conversion to these land uses was based on the suitability of pixels to agricultural and urban land rather than the suitability of natural and semi-natural vegetation. In lulcc, this can most easily be achieved by fitting a binary logistic regression model with no explanatory variables. To do this, a formula such as `Forest~1` should be supplied to the glmModels function.

Methods to evaluate statistical models are provided by the **ROCR** package (Sing et al., 2005), allowing the user to assess model performance using various methods including the receiver operator characteristic (ROC), which is used to measure the performance of models predicting the presence or abscence of a phenomenon (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 fit.

In **lulcc** we extend the native **ROCR** classes to better suit our purposes. The prediction and performance classes of **ROCR** are extended by PredictionList and PerformanceList, respectively, to handle objects of class PredictiveModelList. In the folliwing example we evaluate the logistic regression models using the testing partition from the 1985 observed land use map. Since the Plum Island Ecosystems dataset contains three observed land use maps we could also test the predictive models using data from a subsequent time point. The procedure to evaluate several PredictiveModelList objects using these classes is as follows:

```
> test.data
      <- getPredictiveModelInputData(obs=obs,
                                       ef=ef,
                                       cells=part[["test"]])
> glm.pred    <- PredictionList(models=glm.models,
                                  newdata=test.data)
```

```
   > glm.perf    <- PerformanceList(pred=glm.pred,
                                    measure="rch")
   > rpart.pred <- PredictionList(models=rpart.models,
                                    newdata=test.data)
5  > rpart.perf <- PerformanceList(pred=rpart.pred,
                                    measure="rch")
   > rf.pref     <- PredictionList(models=rf.models,
                                    newdata=test.data)
   > rf.perf     <- PerformanceList(pred=rf.pred,
10                                   measure="rch")
   > plot(list(glm=glm.perf, rpart=rpart.perf, rf=rf.perf))
```

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

15

Another use of ROC analysis is to assess how well the models predict the cells in which gain occurs between two time points. This is only possible if a second observed land use map is available for a subsequent time point. In the following code snippet we perform this type of analysis for the gain of Built between 1985 and 1991. First, we create a data partition

20 in which cells not candidate for gain (cells belonging to Built in 1985) are eliminated. We then assess the ability of the various predictive models to predict the gain of Built in this partition:

```
   > part <- rasterToPoints(obs[[1]],
                            fun=function(x) x != 2,
25                           spatial=TRUE)
   > test.data <- getPredictiveModelInputData(obs=obs,
                                                ef=ef,
                                                cells=part,
                                                t=6)
```

14

```
  > glm.pred  <- Prediction(models=glm.models[[2]],
                            newdata=test.data)
  > glm.perf  <- Performance(pred=glm.pred,
                            measure="rch")
5 > plot(list(glm=glm.perf))
```

Figure 6 shows the resulting ROC curve.

3.4 Demand

Spatially explicit land use change models are normally driven by non-spatial estimates of
10 either the total number of cells occupied by each category at each time point or the number
of transitions among the various categories during each time interval. 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 time point
based on the spatial configuration of the landscape at the previous time point (e.g. Rosa
15 et al., 2013), it is more common to specify the demand for every time point at the beginning
of the simulation (e.g. Pontius and Schneider, 2001; Verburg et al., 2002; Sohl et al., 2007).
In **lulcc** the way in which demand is specified is unique to individual allocation models.
Currently, both allocation models currently included in the package require the total number
of cells belonging to each category at every time point to be supplied as a matrix or
20 data.frame before running the allocation routine.

Land use area may be estimated using non-spatial land use models or, in the case of
a backcast model, national and subnational land use statistics may be used (e.g. Ray and
Pijanowski, 2010; Fuchs et al., 2013). **lulcc** includes a function to interpolate or extrapolate
25 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:

15

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

In reality we are not usually interested in simulating land use change between two time points for which observed land use data is available. However, doing so is useful for model pattern validation, allowing us to test the ability of models to predict the spatial allocation of
5 change given the exact quantity of change.

3.5 Allocation

The allocation algorithm in land use change models determines the pixels in which various land use transitions should take place (Verburg et al., 2002). Currently **lulcc** includes two
10 allocation routines: an implementation of the CLUE-S algorithm and a stochastic ordered procedure based on the algorithm described by Fuchs et al. (2013). Both routines allow the user to optionally provide various decision rules. These are implemented before the main allocation algorithm at each time point and allow the user to incorporate additional knowledge about the study site.

15

3.5.1 Decision rules

The first decision rule included in **lulcc** 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
20 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, such as the transition from shrubland to closed forest (Verburg and Overmars, 2009). The fourth rule prohibits transitions to a certain land use in cells that are not within a user-defined
25 neighbourhood of cells already belonging to that land use. This rule is particularly relevant

16

to cases of deforestation or urbanisation.

Within the `allocate` function the first three decision rules are applied by the `allow` function and the fourth rule is applied by the `allowNeighb` function. For time dependent decision rules the user should supply a land use history raster map, specifying the length of time each pixel has belonged to the current land use. If this is not supplied each pixel is assigned a value of one, representing one model timestep. To apply neighbourhood rules it is necessary to supply corresponding neighbourhood maps to the allocation routine. In **lulcc** these are represented by the `NeighbRasterStack` class. Objects of this class are created with the following command:

```
> w  <- matrix(data=1, nrow=3, ncol=3)
> nb <- NeighbRasterStack(x=obs[[1]], weights=w,
                          categories=c(1,2,3))
```

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 time point.

3.5.2 CLUE-S allocation method

The CLUE-S model implements an iterative procedure to meet the specified demand at each time point and handle competition between land uses. The model is summarised briefly here: for a full description see Verburg et al. (2002) and Castella and Verburg (2007). The algorithm in **lulcc** is based on the description of the model provided by Verburg et al. (2002) only. As a result, for the reasons discussed by Ince et al. (2012), users should not expect to exactly reproduce the output from the original model implementation.

17

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, **lulcc** allows any predictive model supported by PredictiveMod-elList to be used. For each land use the algorithm determines whether the allocated area

5 is less than, equal to or greater than the specified demand. If it is less than or greater than demand the suitability of each pixel in the study region to the land use in question is increased or decreased, respectively, by an amount depending on the difference between the allocated area and specified demand. If the allocated area equals demand the suitability is left unchanged. This procedure is repeated until the demand for all land

10 uses, within a user-defined tolerance, is met. At each iteration the original model perturbs the suitability of each pixel to the various land uses in order to limit the influence of nominal differences in land use suitability on the final model solution. This is replicated in **lulcc** with the parameter `jitter.f`, which controls the upper and lower limits of the uniform random distribution from which the perturbation applied to each pixel is drawn. The default

15 value of `jitter.f` is zero, resulting in a deterministic model. For a full description of the various other parameters supplied to the CLUE-S routine please consult the package documentation.

In **lulcc** allocation models are represented by unique classes. In the following code snip-

20 pet 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 generic `allocate` function:

```
> clues.rules <- matrix(data=1, nrow=3, ncol=3)
> clues.parms <- list(jitter.f=0.0002,
25                     scale.f=0.000001,
                       max.iter=1000,
                       max.diff=50,
                       ave.diff=50)
> clues.model <- CluesModel(obs=obs,
```

```
                                ef=ef,
                                models=glm.models,
                                time=0:14,
                                demand=dmd,
5                               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
10 overcome this limitation we have written the CLUE-S procedure as a C extension using the
.Call interface.

3.5.3 Ordered method

The ordered allocation method is based on the algorithm described by Fuchs et al. (2013).
15 The approach is less computationally expensive and more stable than the CLUE-S algo-
rithm because it doesn't simulate competition between land uses. Instead, land allocation
is performed in a hierarchical way according to the perceived socioeconomic value of
each land use. 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
20 the highest suitability to the current land use are selected for change, where *n* equals the
number of transitions required to meet the demand, as specified by the demand matrix
supplied as an input to the allocation routine. 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.
25 Here, *n* cells with the lowest allocation suitability are converted to a temporary class which
can be allocated to subsequent land uses. The land use with the 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.

We modify the algorithm described by (Fuchs et al., 2013) to allow stochastic transitions. If this option is selected, the allocation 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 allocation 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. To make the model deterministic the user can set the stochastic argument to FALSE when the allocate function is called.

In **lulcc** the ordered allocation 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 allocate function:

```
> ordered.model <- OrderedModel(obs=obs,
                                 ef=ef,
                                 models=glm.models,
                                 time=0:14,
                                 demand=dmd,
                                 order=c(2,1,3))
> ordered.model <- allocate(ordered.model, stochastic=TRUE)
```

3.6 Pattern validation

Spatially explicit land use change models are validated by comparing the initial observed map with an observed and simulated map for a subsequent time point (Pontius et al., 2011). Previous studies have extracted useful information from the three possible two-map comparisons (e.g. Pontius et al., 2008), however, recently Pontius et al. (2011) devised the concept of a three-dimensional contingency table to compare the three maps simu-

20

lataneously. Not only is this approach more parsimonious, it also yields more information about quantity and allocation performance (Pontius et al., 2011). For example, from the table it is straightforward to identify sources of agreement and disagreement considering all land use transitions, all transitions from one land use or a specific transition from one
5 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 useful to perform pattern validation
10 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).

15

In **lulcc**, three-dimensional contingency tables at multiple resolutions are represented by the ThreeMapComparison class. Two subclasses of ThreeMapComparison represent two types of information that can be extracted from the tables: AgreementBudget represents sources of agreement and disagreement between the three maps at several resolutions
20 while FigureOfMerit 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 ThreeMapComparison, AgreementBudget and FigureOfMerit objects allow the user to
25 visualise model performance. The ordered model output for Plum Island Ecosystems is validated in the following way:

```
> ordered.tabs <- ThreeMapComparison(x=ordered.model,
                                   factors=2^(1:8),
                                   timestep=14)
```

21

```
> ordered.agr  <- AgreementBudget(x=ordered.tabs)
> plot(ordered.agr, from=1, to=2)
> ordered.fom  <- FigureOfMerit(x=ordered.tabs)
> plot(ordered.fom, from=1, to=2)
```

5 This procedure was repeated for the CLUE-S model output. The agreement budgets for the transition from Forest to Built for the two allocation procedures are shown by Figure 7, while Figure 8 shows the corresponding figure of merit scores.

4 Discussion

10 The example application for Plum Island Ecosystems demonstrates the key strengths of the **lulcc** 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 programs. A framework in R specifically allows users to take advantage of a wide range of statistical and machine

15 learning techniques for predictive modelling. The framework allows users to experiment with various model structures interactively and provides methods to quickly compare 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.

20

lulcc is substantially different from alternative environmental modelling frameworks. Most importantly, **lulcc** is designed for land use change modelling only, whereas frameworks such as PCRaster and TerraME provide general tools that can be applied to various spatial analysis problems such as land use change, hydrology and ecology. As a result, these

25 tools are targeted towards the model developer rather than the end user. In contrast, most software programs for land use change modelling are designed with the user in mind, with very few providing any way for users or developers to improve or even understand

22

model implementations. With **lulcc** 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 **lulcc** will work across Windows, MacOS and Unix platforms, whereas
5 many existing applications are platform dependent. Comprehensive documentation of the functions, classes and methods of **lulcc**, 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.

10 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 (Hijmans, 2014).
15 **lulcc** 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 offers one possible solution to this problem, allowing predictive models to be used in a memory-safe way. In
20 effect, this would mean spatially explicit input data including observed land use maps and explanatory 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 **lulcc** because it is excessively time consuming compared to the current approach. Despite this limitation, since most applications involve a relatively small geographic extent
25 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 **lulcc** applications to fail. For example, the CluesModel and OrderedModel objects from the above example each had a size of approximately 40Mb, which is easily handled by modern personal computers. On a 64-bit machine with Intel Core i3 @ 1.4 GHz and 4Gb RAM, the allocation methods

for the two Model objects took 50 seconds and 8 seconds, respectively.

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 time point 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 allocation of previous deforestation events, updated after the first round. One advantage of this approach is that it accounts for uncertainty in the quantity and allocation of change simultaneously, whereas the current routines in **lulcc** only consider the allocation 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 **lulcc** with this class of model would be a valuable addition to the software because land use change is increasingly recognised as an issue with drivers and implications at local, regional, continental and global levels.

An important contribution of **lulcc** is to provide modules to assist with model pattern validation, a crucial aspect of model development that is nevertheless frequently overlooked within the land use change modelling community (Rosa et al., 2014). A further improvement that could be made to the package is to incorporate more sophisticated ways of fitting and testing the predictive models that estimate land use suitability. For example, a routine to calculate the Total Operating Characteristic (TOC) (Pontius and Parmentier, 2014) would improve upon the ROC analysis currently supported. While ROC shows two ratios, hits/(hits+misses) and false alarms/(false alarms+correct rejections), at multiple resolutions, TOC reveals the quantities used to calculate these ratios, allowing greater

24

interpretation of model diagnostic ability.

One of the main strengths of **lulcc** 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 **lulcc** to broaden the available model structures and, therefore, improve the ability of **lulcc** to capture model structural uncertainty. The methods in the current version of **lulcc** 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). This means that, unlike inductive models, they can be used to establish causality between land use change and its driving factors (Overmars et al., 2007). Including this class of model in **lulcc** would allow inductive and deductive land use change models with different spatial resolutions to be dynamically coupled in order to better capture the complexity of the land use system (Moreira et al., 2009).

Free and open source software improves 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 **lulcc**. 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 relatively straightforward task for commonly used compiled languages (C/C++, Fortran). Users may also develop their own R packages that depend on **lulcc** 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

25

explanatory variables) for inclusion in the package.

5 Conclusions

In this paper we have presented **lulcc**, a free and open source software package providing an object-oriented framework for land use change modelling in R. **lulcc** allows various aspects of the modelling process to be performed within the same environment, supports three types of predictive model and includes two allocation routines. The modelling process can be expressed programmatically, facilitating reproducible science. 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 can identify improvements to the code and, under the terms of the licence, are free to redistribute changes to the wider community. We view **lulcc** 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 R project for statistical computing is available for Windows, MacOS and several Unix platforms. To download R, visit the project homepage: https://www.r-project.org/. Two popular and free integrated development environments (IDEs) are provided by RStudio (https://www.rstudio.com/) and ESS (http://ess.r-project.org/). We suggest that potential **lulcc** users familiarise themselves with the **raster** package by reading the "Introduction to the raster package" vignette, available on the package homepage: https://cran.r-project.org/web/packages/raster/.

The **lulcc** source code currently resides on CRAN. This paper corresponds to version 1.0 of the package. It can be downloaded from the R command line as follows:

```
> install.packages("lulcc")
```

The script for the Plum Island Ecosystems application is available as a demo within the package. To load the package and run the demo, type the following commands:

```
> library(lulcc)
> demo(package = "lulcc")
> demo(topic = "gmd-paper")
```

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.

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.

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.

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.

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.

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.

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.

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

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.

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.

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.

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

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

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

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.

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.

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.

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.

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.

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

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.

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.

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.

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.

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.

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.

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.

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

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.

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.

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.

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.

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.

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.

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.

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.

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.

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.

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.

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

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.

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.

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.

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.

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.

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.

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.

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.

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

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.

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

31

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 inter-
5 comparison study, Geophys. Res. Lett., 36, L14814, doi:10.1029/2009GL039076, 2009.

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.

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.

10 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.

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
15 land-change models, Ecol. Model., 179, 445–461, doi:10.1016/j.ecolmodel.2004.05.010, 2004a.

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.

Pontius, R. G., Boersma, W., Castella, J., Clarke, K., Nijs, T., Dietzel, C., Duan, Z., Fotsing, E.,
20 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.

Pontius, R. G., Peethambaram, S., and Castella, J.: Comparison of three maps at multiple reso-
25 lutions: 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.

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
30 and uncertainties from a global multimodel ensemble experiment, P. Natl. Acad. Sci. USA, 111, 3262–3267, doi:10.1073/pnas.1222473110, 2014.

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.

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.

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.

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.

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.

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.

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.

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.

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.

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

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.

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.

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.

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.

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.

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.

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.

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.

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.

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.

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.

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.

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.

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.

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.

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

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.

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.

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.

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.

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.

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.

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.

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.

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.

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.
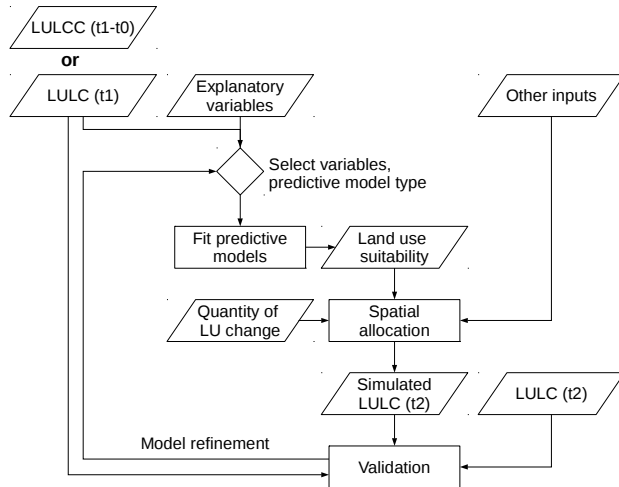
Figure 1. Diagram showing the general methodology used for inductive land use change modelling applications, adapted from Mas et al. (2014). The input land use/land cover data can be a single categorical map showing the pattern of land use/land cover at one time point (LULC (t1)) or a series of maps showing historical land use/land cover transitions (LULCC (t1-t0)).
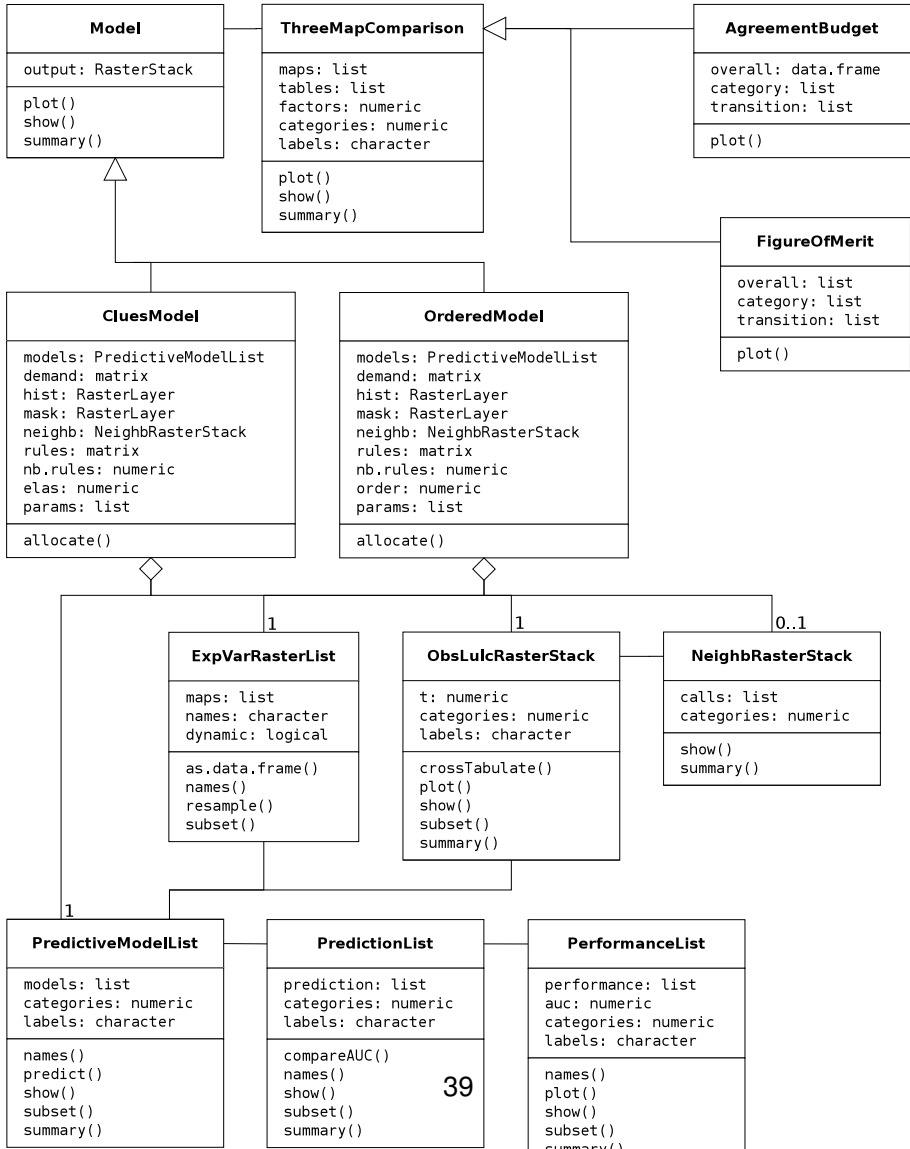
Discussion Paper | Discussion Paper | Discussion Paper | Discussion Paper |

Model

output: RasterStack

plot()
show()
summary()

ThreeMapComparison

maps: list
tables: list
factors: numeric
categories: numeric
labels: character

plot()
show()
summary()

AgreementBudget

overall: data.frame
category: list
transition: list

plot()

FigureOfMerit

overall: list
category: list
transition: list

plot()

CluesModel

models: PredictiveModelList
demand: matrix
hist: RasterLayer
mask: RasterLayer
neighb: NeighbRasterStack
rules: matrix
nb.rules: numeric
elas: numeric
params: list

allocate()

OrderedModel

models: PredictiveModelList
demand: matrix
hist: RasterLayer
mask: RasterLayer
neighb: NeighbRasterStack
rules: matrix
nb.rules: numeric
order: numeric
params: list

allocate()

1

ExpVarRasterList

maps: list
names: character
dynamic: logical

as.data.frame()
names()
resample()
subset()

1

ObsLulcRasterStack

t: numeric
categories: numeric
labels: character

crossTabulate()
plot()
show()
subset()
summary()

0..1

NeighbRasterStack

calls: list
categories: numeric

show()
summary()

1

PredictiveModelList

models: list
categories: numeric
labels: character

names()
predict()
show()
subset()
summary()

PredictionList

prediction: list
categories: numeric
labels: character

compareAUC()
names()
show()
subset()
summary()

PerformanceList

performance: list
auc: numeric
categories: numeric
labels: character

names()
plot()
show()
subset()
summary()

39

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

Table 1. Functions included in the **lulcc** package

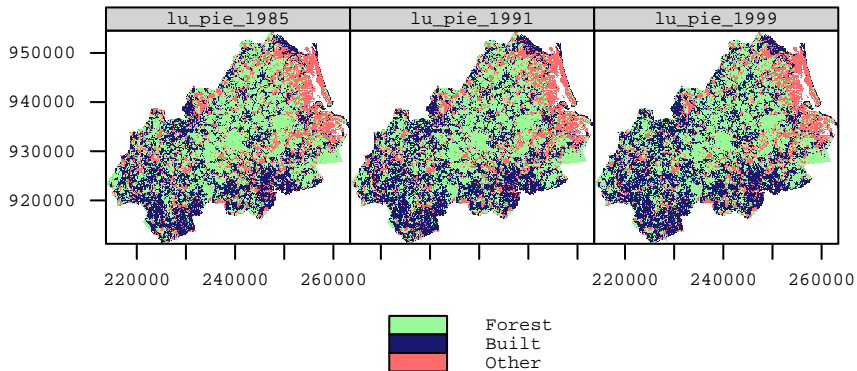| Function name	Description
AgreementBudget	Calculate agreement budget (Pontius et al., 2011)
getPredictiveModelInputData	Create data.frame with variables required to fit predictive models
allocate	Perform spatial allocation using various methods
approxExtrapDemand	Create a demand scenario by linear extrapolation
compareAUC	Compare the area under the curve (AUC) for various predictive models
crossTabulate	Calculate the contingency table for two categorical raster maps
FigureOfMerit	Calculate the figure of merit (Pontius et al., 2011)
glmModels	Fit multiple glm models
NeighbRasterStack	Calculate neighbourhood values
partition	Partition Raster* map
PredictionList	Create a ROCR prediction object for each model in a PredictiveModelList object
PerformanceList	Create a ROCR performance object for each prediction object contained in a PredictionList object
predict	Make predictions using a PredictiveModelList object
randomForestModels	Fit multiple random forest models
rpartModels	Fit multiple recursive partitioning and regression tree models
resample	Resample an ExpVarRasterList object to the parameters of an ObsLulcRasterStack object
ThreeMapComparison	Calculate three-dimensional contingency tables (Pontius et al., 2011)
total	Sum the total number of cells belonging to each class of a categorical raster map

Figure 3. Observed land use maps for the Plum Island Ecosystems site in 1985, 1991 and 1999, created by plotting the ObsLulcRasterStack object representing the data.
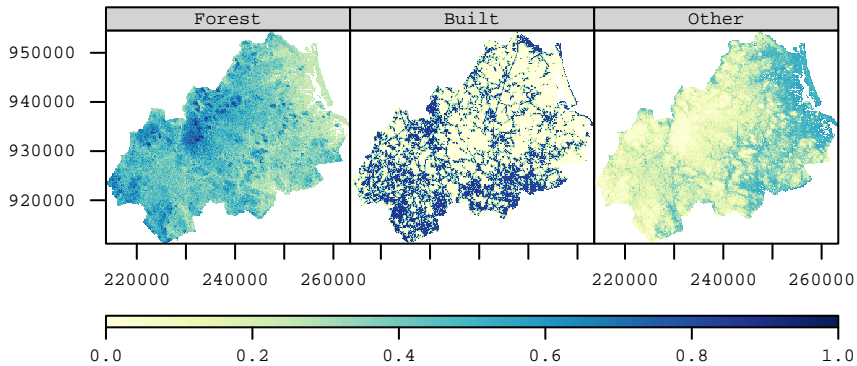
Figure 4. Suitability of pixels in the Plum Island Ecosystems study site to belong to Forest, Built and Other land use classes according to binary logistic regression models. Elevation and slope are used as explanatory variables for all land uses while Built additionally includes distance to built pixels in 1985.
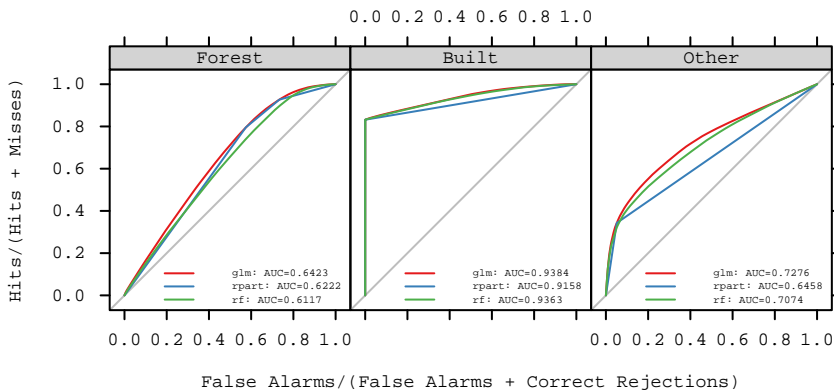
Figure 5. ROC curves showing the ability of each type of predictive model to simulate the observed pattern of land use in the Plum Island Ecosystems site in 1985 in the data partition left out of the fitting procedure.
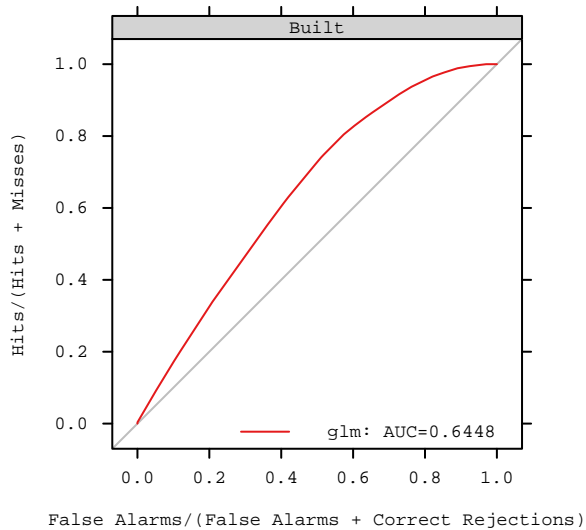
44

Figure 6. ROC curve showing the ability of the binary logistic regression model fitted on observed land use data from 1985 to predict the gain in Built land between 1985 and 1991.
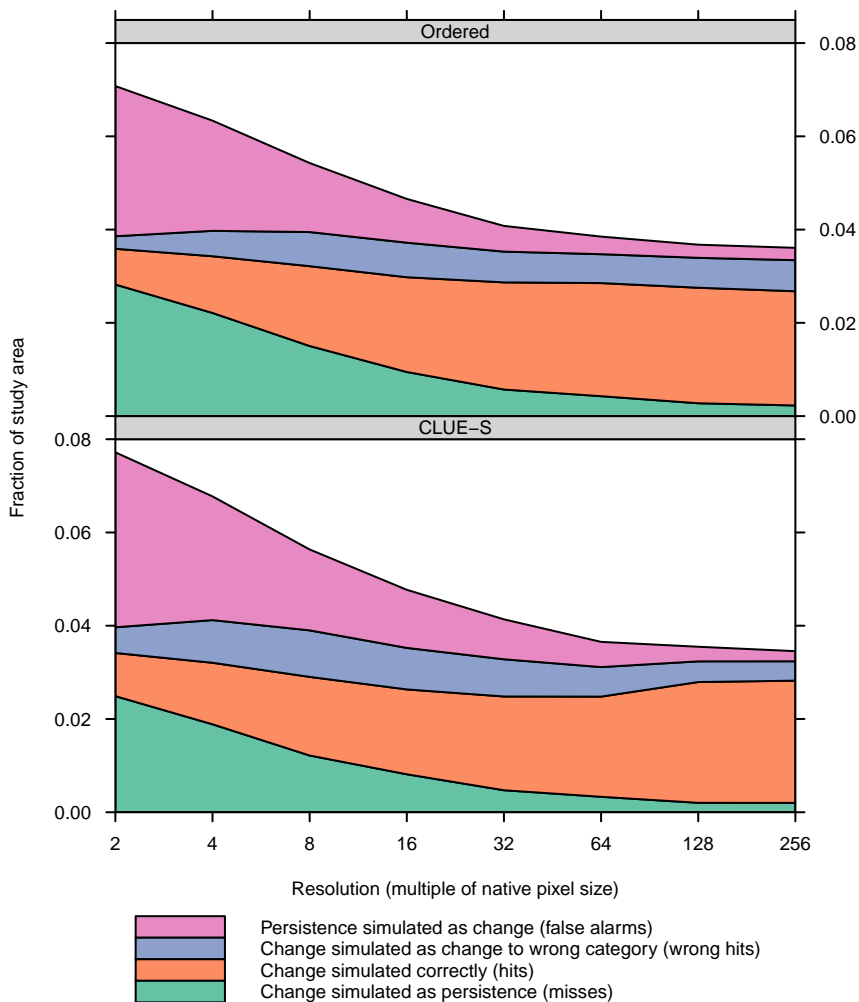
47

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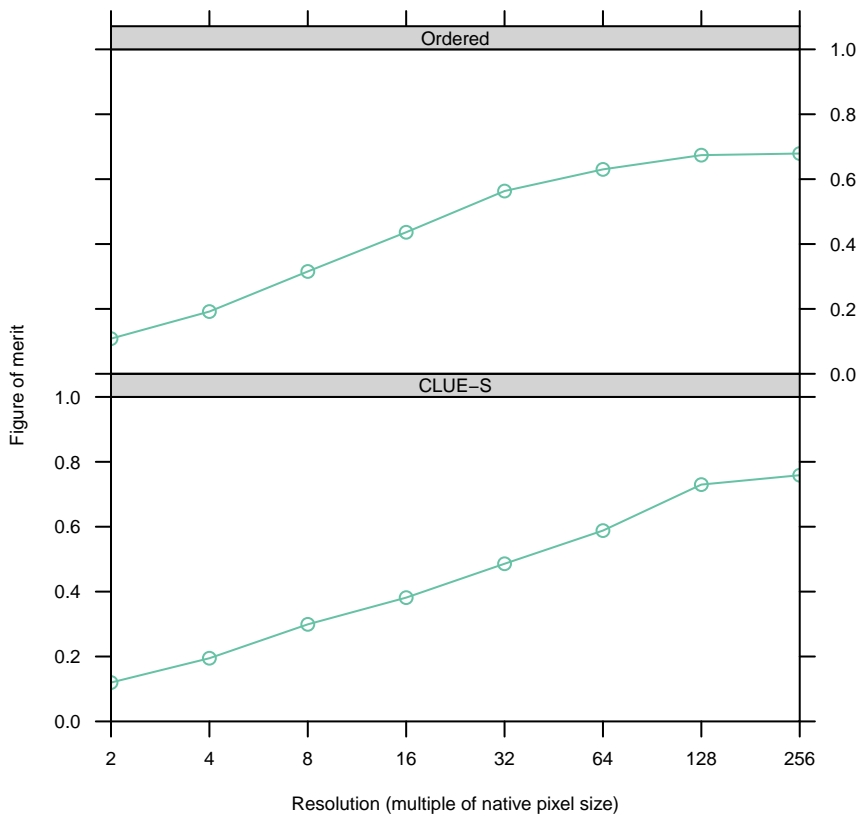
Figure 8. Figure of merit scores corresponding to the agreement budgets depicted in Figure 7.