Author comments

Simon Moulds

July 30, 2015

We thank both referees for their thoughtful comments on the manuscript and software. Here, we address each reviewer comment in blue and, where appropriate, detail changes in the revised manuscript in magenta.

Response to Referee 1 comments:

1. I am an appropriate reviewer because this manuscript is directly in my area of expertise, which is spatially explicit land change modeling. The authors have done an impressive amount of computer programming. The research community will benefit from at least some, if not all, of the various modules in R. There were numerous passages of the submitted manuscript that I found very confusing. Also the manuscript is more difficult than necessary and longer than necessary for a variety of reasons. Below I make suggestions for improvements. I hope a major revision to the manuscript can render the manuscript publishable.

   First, the authors must cut all non-essential information. The manuscript is too long. The reader becomes exhausted, therefore can miss some important points. I needed three separate sittings to slug through the manuscript. The main point of the manuscript is to describe the software. Anyone who will be interested in reading this paper already knows that land change is important for a variety of reasons. Therefore the Abstract should not have sentences such as "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". The authors should cut the first four sentences of the Introduction. The Introduction could begin with "Land use change models are...". The authors should cut the entire second paragraph of the Introduction. Maybe the manuscript should begin with "Spatially explicit land use change models are commonly written in...". The statement of the main purpose of the manuscript should be in the first paragraph of the Introduction. This manuscript's target audience is technically oriented people who might use the R code. The manuscript must focus on that particular audience. This manuscript does not need to explain why land change is important. Moreover, the manuscript does not even need to explain why modeling is important. The manuscript must focus on describing why land change modelers might want to use the authors' software.

This constructive criticism of the manuscript by the reviewer has made it clear that our original manuscript lacked focus and did not provide a good advertisement for our software. The revised manuscript is much shorter, and focuses solely on describing the R package and explaining why users may want to use the modules it contains. The sentence in the Abstract referred to by the reviewer has been removed. The entire second paragraph and much of the first paragraph of the Introduction has also been removed. In the revised manuscript, the first paragraph states the purpose of the manuscript:

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

2. The manuscript frequently uses the word "different" where the word "various" would be more precise. For example, "Detailed reviews of different models and modelling approaches are available..." is more clearly stated as "Detailed reviews of various models and modelling approaches are available...". The word "different" makes the reader wonder "different than what?"

   We have reconsidered our use of the word "different" and replaced it with more precise alternatives where appropriate.

3. Please use the word "allocation" rather than "location" throughout the manuscript for the reasons stated in Pontius et al. (2011).

   For every occurrence of "location" we have either replaced it with "allocation" or rewritten the sentence. We thank the reviewer for pointing out this lack of precision.

4. The manuscript should avoid using the word "scale" because that word means too many different things. For example the manuscript says "an earlier version of CLUE-S that operates at larger spatial scales". Does scale mean extent or resolution, and if so what does larger mean. I think the answer is neither extent nor resolution. I think first CLUE allowed pixels that have partial membership to multiple categories, but then CLUE-S assumes each pixel has full membership to exactly one category. Those types of category memberships are not necessarily related to extent or resolution.

   We agree that the use of the word "scale" is potentially confusing and we avoid it in the revised manuscript. Regarding the example the reviewer mentions about CLUE/CLUE-S, we were attempting to make the point that the original CLUE was designed to work at national and continental levels (Verburg et al., 2002). we agree that the sentence is not clear and have therefore rewritten it as follows:

"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."

5. Please cut the word "every" from line 5 of page 3365. Various approaches have various stages, many of which are not covered by the software's paradigm. For example, the user interface has no place for discussions with stakeholders in order to develop scenario storylines, which are crucial for some modelling approaches.

   The reviewer is correct to point out this mistake: we did not intend to claim our software could perform every stage of all land use change modelling approaches. Instead, we meant that our software could perform the stages outlined in Figure 1 of the manuscript, which depicts an idealised process for an inductive, spatially explicit modelling approach. However, we acknowledge that even this claim could be misleading, so we have rewritten the sentence as follows:

   *"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."*

6. Scientific manuscripts should use the word "significant" if and only if the word means that a p-value is less than the alpha-level in a statistical hypothesis test. Please replace uses of the word "significant" unless they refer to inferential statistics.

   We have deleted the sentence to which one use of the word "significant" belonged, and replaced the other occurrence with the word "importantly".

7. I have no idea the meaning of the sentence in line 1 of page 3366.

   The sentence to which the reviewer refers, "Significantly, given the importance of predictive models to land use change modelling, R has become the standard development platform for statistical software", was supposed to convey that one important reason the R language is useful for land use change modellers is that they automatically have access to a wide range of statistical tools (such as Random Forest classification) that may be useful for land use change modelling. To be clearer, we have removed the sentence and rewritten the preceding sentence as follows:

   "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."

8. Authors should cut most of the description of the study sites. Readers wonder why it is important to know about hydrology in the Plum Island Ecosystems (PIE), then readers realize that hydrology is irrelevant to the manuscript's purpose. Thus readers become more exhausted and districted. The manuscript forces the reader to constantly make judgements between which sentences are important and which sentences are

not important. For example, it is not important that a map for 2005 for PIE cannot be used, nevertheless the manuscript refers to this unused map of 2005. The authors must simply describe the data that they actually analyzed. The manuscript must stick to its one point, which is to describe the application of the authors' modules in R. It is not clear why two case studies are needed. If the concepts are the same in the software for all case studies, then [one] example application should suffice. Two case studies would only be necessary only if the two cases had different data formats, such as raster versus vector. However, for the two case studies of Plum Island Ecosystems and Sibuyan, the second case study seems to give no additional insight concerning the R software.

We apologise for the lack of clarity regarding the two example datasets included with the package. We have taken the advice of the reviewer and rewritten Section 3.1 more concisely. It now reads:

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

9. Section 3.2 must state clearly whether the R modules assume that each pixel belongs completely to exactly one category, meaning mixed pixels are not allowed.

   We have clarified this by appending the following sentence to the first paragraph of Section 3.2:

   "The current version of the software only supports categorical land use data, which means that each pixel must belong to exactly one category."

10. The use of the word "timestep" on page 3369 is very confusing, because "timestep" means the duration between two time points. I think "timestep" should be "time point".

    The reviewer is correct: in this case "timestep" should be "time point". We have corrected this mistake throughout the manuscript.

11. The meaning of "correct spatial resolution" is not clear on page 3369.

    We agree that this phrase is confusing. We meant that additional input maps should have the same spatial resolution and extent as the ObsLulcRasterStack object that defines the study region. We have removed the sentence to which the reviewer refers, and added the following text to the end of Section 3.2:

    "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."*

12. Section 3.3 should begin with the sentence "Inductive land use change models relate the...". The second paragraph of section 3.3 should be "Parametric models, such as logistic regression, assume the error terms of the input data to be..."

    We have changed these sentences in the revised manuscript. The first sentence of Section 3.3 is now written:

    Inductive land use change models relate the pattern of observed land use to spatially explicit explanatory variables."

    The second paragraph of Section 3.3 now begins:

    Parametric models, such as logistic regression, assume the error terms of the input data to be independent and identically distributed (Overmars et al., 2003; Wu et al., 2009).

13. The authors should cut all information in section 3.3 that does not relate to the R modules, for example the discussion of non-parametric models.

    We have removed most of this discussion as well as the background information on the receiver operator characteristic in a later paragraph. We believe it is worth mentioning that, even though non-parametric models do not make the same assumptions as parametric modes, they may still be affected by spatial autocorrelation. In the revised manuscript this is only briefly mentioned:

    "Parametric models, such as logistic regression, assume the error terms of 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). 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."

14. In line 5 of page 3372, should "occurrence" be "gain"?

Yes, it should. The particular sentence to which the reviewer refers has been cut from the revised manuscript, but we have clarified our use of "occurrence" and "gain" in other parts of the manuscript.

15. It is extremely confusing to the term "null model" in line 9 of page 3372 because "null" means a prediction of complete persistence in much of the other literature in land change modelling. I am very confused by figure 4 and the sentence: "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." It seems to me that there should be one suitability map for the gain of each category. It is possible for Forest to gain, and for Built to gain, and for Other to gain; so it seems there should be three suitability maps, one for Forest gain, one for Built gain, and one for Other gain. Any gain implies a loss of some other category, depending on where the gain occurs.

    We realise that the use of the term "null model" may be confusing for readers who have encountered the word "null" to describe a prediction of complete persistence. Indeed, we realise the concept of a "null model" may be confusing. In this case, we decided to use a null model following Verburg and Overmars (2009): "In case of (semi-) natural land use types a uniform suitability is assigned to all locations because the selection of locations for reclamation or abandonment is based on the suitability for the agricultural or urban use rather than on the suitability for (semi-) natural vegetation". The reviewers comments have made us realise that such an approach may be unnecessarily simplistic and confusing to readers. In the revised manuscript, therefore, we have included a suitability map for Forest based on elevation and slope. We maintain, however, that in some circumstances a model without explanatory variables may be appropriate for some models and/or case studies, and we have therefore made the following addition to Section 3.3:

    "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."*

16. I think "plot" should be "map" in lines 15 and 17 of page 3372. I think "model" should be "fit" in line 30 of 3372.

    We agree with the reviewer. Section 3.3 has changed substantially in the revised manuscript. The sentences to which the reviewer refers is now written:

"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 each land use over the study region. 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)..."*

We agree that "model" on line 30 should be "fit", and have replaced this in the revised manuscript. The sentence is now written:

"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."

17. It would be much better for the software to use the Total Operating Characteristic (TOC) rather than the Relative Operating Characteristic (ROC), for reasons explained by Pontius and Parmentier (2014). My students have created R code for TOC.

    We agree that the TOC would be a useful addition to the software but would also argue that the ROC remains a useful, if not perfect, measure of model performance. We are pleased that the reviewers students have created R code to calculate the TOC. It would seem pointless for us to repeat their work, especially since the TOC has applications beyond land use change modelling, so our suggestion would be for the reviewers students to write their TOC code as an R package (perhaps following the structure of ROCR (Sing et al., 2005)). A future version of lulcc could then utilise its functionality. An alternative approach would be for the reviewers students to collaborate with us to incorporate their TOC code into our package, in which case they would become package contributors. As far as the manuscript is concerned, we have added the following text to the Discussion in order to highlight the fact that the ability to calculate and plot the TOC would improve upon the existing functionality:

    "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 interpretation of model diagnostic ability."

18. In section 3.4, the word "timestep" is again potentially confusing. Section 3.4 must distinguish between the specification of the area of each category versus the specification of the area of each transition among all the various categories. For example, in Idrisi's Land Change Modeler, the user must specify a Markov transition matrix that determines the sizes of the transitions; the user does not enter the size of the area of each category.

Please see our response to comment 10 regarding the use of "timestep" and "time point". We are grateful to the reviewer for highlighting the various ways of specifiying demand. In the revised manuscript we have altered the first paragraph of Section 3.4 as follows:

"Spatially explicit land use change models are normally driven by non-spatial estimates of either the total number of cells occupied by each category or the number of transitions among the various categories at each time point. 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 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 data.frame before running the allocation routine."*

19. I was confused by lines 21-24 on page 3374. If those lines are not essential, then one approach is to cut them.

    We agree with the reviewer that these lines ("While the set of included in lulccR could be used as the basis of a simple agent-based model...") are not essential for the description of the software and therefore potentially confusing. The passage has been removed from the revised manuscript.

20. Section 3.5.1 must discuss how the algorithms deal with competition, for example in PIE, both Built and Other can compete to gain from Forest. If a Forest pixel has large suitabilities for both Built and Forest, then how does the software decide whether built or Other gains from the Forest pixel.

    Section 3.5.1 only explains the various decision rules that can be supplied to the two allocation functions. Sections 3.5.2 and 3.5.3 explain each routine in detail, including how each deals with competition between land uses. We have emphasised this in the revised manuscript, so that the first paragraph of Section 3.5 reads:

    "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 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."*

21. I do not know the specific meaning of "comparable" in lines 26 and 28 of page 3377. Please clarify, because anything can be compared.

    This comment and the one following have made it clear that these sentences are confusing. It was our intention to compare the predicted land use maps for Sibuyan Island from our implementation of CLUE-S with the predicted maps from the original CLUE-S (as presented by Verburg et al. (2002)). However, we now think that there are too many unknowns regarding the model setup of the original CLUE-S for this to be a valid comparison. We have therefore cut lines 24-28 of page 3377, and emphasised in the beginning of section 3.5.2 that our implementation is based on the algorithms as they are described by Verburg et al. (2002), rather than an attempt to replicate exactly the original model implementation:

    *"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."*

22. Further explanation is required for the sentence "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." I do not even know the meaning of the "original model" and "this model". The entire manuscript concerns the model interface, so this seems to be an important limitation that must be described in depth.

    We agree that this sentence is unclear. By "original model interface" we meant the interface for the original version of CLUE-S (i.e. the version that can be downloaded here: http://www.ivm.vu.nl/en/Organisation/departments/spatial-analysis-decision-support/Clue/) that has restrictions concerning system requirements (e.g. 32-bit Windows only) and number of cells included in the analysis. In any case, please see our response to the previous comment.

23. Section 3.5.3 should make it clear that the suitability maps can influence the size of each transition from one category to another category.

    We are unsure what the reviewer means here: the suitability maps do not affect the number of simulated transitions, since this is governed by the demand matrix which is supplied as an input to the model. For example, assuming Built has the highest socioeconomic value of any land uses in a given study region, if the number of cells allocated to Built at time 0 is 1000, and the demand for Built at time 2 is 1500, then 500 cells that have the highest Built suitability but which don't already belong to Built are changed. The confusion may arise from lines 10-12 on page 3378, which we have rewritten as follows:

    "In this case, n cells with the highest suitability for 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."

24. Section 3.5.3 describes how the authors modified the algorithm to allow for stochastic transition. I cringe when models have stochastic components, because then each run is different, thus debugging and interpretation become much more complicated than they would otherwise be. There seems to be several points where the authors inserted stochastic components into the R code. These stochastic components are one reason why I might not use some modules of the R code.

    We appreciate the reviewers reservations about stochastic modelling but would argue that in some applications this feature may be useful to capture some of the uncertainty inherent to land use change modelling, particularly when the model output is passed to other models such as hydrological models.

    We have inserted stochastic components into the two allocation routines described in Sections 3.5.2 and 3.5.3. Our implementation of CLUE-S is stochastic because we include the parameter `jitter.f`, which perturbs the land use suitability at each time. The original CLUE-S implementation contains a similar parameter except users are only permitted to vary it within a small range, and the documentation does not make clear that the number controls the degree of pertubation of the land use suitability at each time. In our view it is better to explain to the user what the parameter does and give them reasonable control over its value. The default value of `jitter.f` is 0, which means the model is not stochastic. We have changed lines 5-6 of page 3377 to be more clear about this point:

    *"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 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."*

    Again, in the ordered allocation procedure the user can turn off stochastic transitions by setting `stochastic=FALSE` when the allocate routine is called. We have added the following sentence to the first paragraph of Section 3.5.3:

    "To make the model deterministic the user can set the `stochastic` *argument to FALSE when the* `allocate` *function is called."*

25. The title of section 3.6 should be "Pattern Validation" rather than "Validation" to distinguish from Process Validation.

    We have changed the title of section 3.6 as the reviewer suggests.

26. In Section 3.6, "Pontius et al. 2007" should be "Pontius et al. 2008".

    We apologise for this mistake, which has been corrected in the revised manuscript.

27. Line 14 of page 3379 should change from "allocation performance" to "quantity and allocation performance".

    We have made this correction. The sentence now reads:

    "Not only is this approach more parsimonious, it also yields more information about quantity and allocation performance (Pontius et al., 2011)."

28. Line 22 of page 3379 should change from "common" to "useful". In fact, it is not common, but hopefully your software will make it more common.

    We have made this correction. Indeed, we hope so too.

29. The authors should add the criterion of "well documented" to line 22. If the algorithms are not well documented, then freely available software is useless. Poor documentation is the number constraint to advancement of the science of land change modeling.

    We agree with the reviewer that well documented code is essential but we're not clear to which page the reviewer refers with this comment. Perhaps the best place to add this criterion is in the Design goals. Paragraph 3 of this section now contains the sentence:

    "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."

30. Why do I not see any years listed in the citations?

    The citation style used by Copernicus journals places the year of publication at the end, or near the end, of the citation.

31. Figure 4 must say the suitability for what?

    We have rewritten the caption for Figure 4 as follows:

    "Suitability of pixels in the Plum Island Ecosystems study site 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."

32. Figure 3 should please follow the recommendations of Pontius and Parmentier (2014). Most importantly, the software must allow for a mask to eliminate pixels that are not candidate for gain. For example, if you are simulating the gain of Built beyond time 1, then all pixels that are in a Built state at time 1 are not candidates for gain of Built beyond time 1, so these pixels must be eliminated from the ROC analysis. The shape of the curve for Built in figure 3 makes me believe that the authors did not eliminate those pixels. This is a common blunder in the profession.

Figure 3 shows the ability of the various predictive models to predict the location of Forest, Built and Other land uses in cells belonging to the testing partition (in this case all the cells that were not used to fit the models) at the same time point. The figure is not showing the ability of the models to predict the allocation of gain, so, as far as we understand, it wouldn't make sense to remove pixels that are not candidate for gain. However, we acknowledge that the ROC curve may be used in this way and have therefore added an additional paragraph and figure to Section 3.3, as follows:

"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 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..."

The resulting ROC curve has an AUC value of 0.6448, which is possibly closer to what the reviewer was expecting. We thank the reviewer for making this observation: using ROC in the way he suggests is an important use case of our software.

33. Figure 3 also needs axis labels. The vertical axis should have the label ("Hits/(Hits + Misses)" and the horizontal axis should be "False Alarms/(False Alarms + Correct Rejections)". In any case, it would be better to show TOC plots, rather than ROC plots.

    We have added the labels the reviewer suggests. Regarding the use of TOC, please see response to comment 17.

34. The vertical axis for figure 6 should range from 0 to 0.16, so readers can see the crucial regions of the figure. Also, in the legend for figure 6 have the words: "Misses", "Hits", "Wrong hits", "False alarms", and "Correct rejections" from bottom to top to accompany the longer descriptions. It is helpful to have one-word or two-word descriptors to refer to those categories. I thank the authors for writing R code to compute figure 6. I hope many readers will use the authors' R module to perform pattern validation similar in format to figure 6. This is an important contribution.

    We have adjusted the plot method so that the default legend includes the short descriptors. Figure 6 has in fact now been removed, for the reasons discussed in our response to comment 21. The agreement budget plot for the specific transition Forest to Built for the two allocation modules is shown by Figure 7.

35. The vertical axis labels on figures 7 and 8 are extremely alienating. There are many missing numbers. It seems the left axis should have numbers to describe to full range. I do not see any need for numbers on the right axis.

We have corrected the plot method to give more helpful axis labels.

36. Wow, this review process has been exhausting for me. I committed the energy and many hours because the authors are doing important work. I hope my feedback helps.

    We thank the reviewer for committing the time and energy to review our manuscript. In our view his comments have greatly improved both the manuscript and the software.

Response to Referee 2 comments:

1. I congratulate the authors with a very readable paper on land use change modelling and how this can be done transparently and reproducibly with a package they developed for R. Not only does it describe and introduce the software well, it also gives a very extensive literature review to modern environmental modelling paradigms as well as land use change modelling approaches. As I am not an expert in the area of land use change modelling, in this review I will focus on the software development side, and whether the paper reaches the aim of empowering land use change modellers, and inviting them to take the modelling process in their own hands.

   We thank the reviewer for his feedback on this aspect of the paper: his expertise has improved the manuscript and software enormously.

2. For a paper introducing a software framework, it is extremely extensive on describing land use change modelling, but extremely thin on describing the software. The class diagram is offered in a UML diagram that gives only the main classes and core functions; methods are not even mentioned. Will the land use change modeller be helped by this, and be invited to understand it, use it and extend it? The least the authors need to do is explaining the arrow types and symbols in the UML. A table with all the methods and key function offered would also be helpful, as these are the things a user will need first.

   Bearing in mind the other changes to the software that the reviewer has suggested we have redrawn the UML diagram. It now includes a complete list of operations (methods) defined for each class in the package as well as a key to help readers understand the structure of the package. We have included a table with the main functions included in the package. We thank the reviewer for this suggestion.

3. The paper also needs to be much more explicit about which users it wants to attract, and serve. Should the package users be fluent with the packages sp, raster, caret, rgdal and maybe more? Or should more novice R users also feel invited? The current examples, which should be the package advertisement, contain constructs like `raster::extract`, `obs@maps[[1]]`, save a plot to `p` to later plot it with `print(p)` - all constructs that will scare novice users, and that may not be necessary.

We thank the reviewer for this observation: it is an important criticism since one of the main objectives of the software package is to be accessible to R users of all abilities who may not yet be familiar with all of the packages upon which **lulcc** depends. We have therefore made changes to both the manuscript, by removing potentially confusing constructs and providing better descriptions of package dependencies, as well as the code, by ensuring that it can be expressed in a more approachable way. For example, we have now written indexing methods so that the user does not have to directly access S4 object slots (e.g. now, users can simply write `obs[[1]]` instead of `obs@maps[[1]]`). Elsewhere, we have removed references to **rgdal** and **caret** packages to avoid the implication that prior knowledge of these packages is necessary. Further, in the Code availability section we have added the suggestion that potential users consult the "Introduction to the raster package" vignette, since **lulcc** relies heavily on **raster**.

4. After the first action (p11, l11), I was surprised that (i) `pie` is a list that is available in the package (and not e.g. a data object loaded by `data(pie)`, (ii) that the object created is nothing but a RasterStack, the categories, their names, and a set of times. Why not create a RasterStack that holds all this information? In that case, `ObsLulcMaps` could simply extend `RasterStack`, and would get all its methods (like `plot`!!) for free.

   The two datasets included with the package have now been saved as R data objects that can be loaded by `data(pie)`. For the time being the data objects are still lists (not least because this is the case in the version submitted to CRAN) but perhaps in future versions we will change this. We thank the reviewer for this suggestion, because it makes the package easier to understand.

   The second suggestion is indeed a more elegant and useful way of formulating the class and we have changed the definition of `ObsLulcMaps` and `NeighbMaps` classes accordingly. Note that the `ExpVarMaps` class cannot be treated as a RasterStack object because it allows raster maps with different extents and resolutions. On a related matter, we have thought more deeply about the class names and the information they convey. We have decided that the name should indicate the main data strucure represented by the class. For example, the ObsLulcMaps class has been renamed ObsLulcRasterStack, while the ExpVarMaps class has been renamed ExpVarRasterList. We hope this makes the classes easier to understand.

5. In figure 1, it is not clear from the caption what t0, t1 and t2 refer to, and why either LULCC (t1-t0) or LULC (t1) can be input.

   We have changed the caption of Figure 1 to the following:

   "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)).”

6. I am not very fond of R packages with an R in them. In this case, it leads to expressions like “the lulccR R package”, “the lulccR package for R”, or ”the lulccR package [...] written in the R programming language”. (abstract), which are all odd. I would suggest to rename it into `lulcc`, and start sentences now starting with *lulccR* with “The *lulcc* R package...”

   We have renamed the package to `lulcc` as suggested and changed the awkward sentence structures the reviewer points out.

7. The package is currently available on GitHub. Why has it not been submitted to CRAN? Submitting to CRAN offers easier accessibility (and allows to remove the now very odd lines 9-10 on page 11), quality control, and archiving by a third party: the current version 0.1.0 may change at any moment. Also, a version number should be removed from the title, but it would be appropriate to have the paper correspond to a 1.0 version on CRAN, indicating the author’s opinion that it is mature enough to be published. Did the package get any use by others that the authors can report on e.g. in papers published?

   We have now (30/07/2015) submitted version 1.0 of the package to CRAN. So far the package has been used internally and, as far as we know, only by one or two others around the world. We hope this will increase, especially when the package is available on CRAN. In anticipation of passing the various CRAN checks, we have added the following text to the Code availability section:

   *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...*

8. The classes provided by lulccR seem to be useful, but why does the package not come with methods that users expect? After creating `obs`, I tried `plot(obs)`, `summary(obs)`, but none of them worked. Now, users not only need to type the long `AgreementBudget.plot` and `FigureOfMerit.plot`, but they also need to memorize it, instead of simply using `plot`. This needs to be simplified; similar to saving the plot, then printing it, or only saving it (p22, l17,19).

   We thank the reviewer for drawing our attention to these deficiencies. We have now written `plot`, `summary` and `show` methods for the S4 objects in the package and removed confusing and long function names such as `AgreementBudget.plot`. In addition, we have included coercion methods (for example, from `ObsLulcRasterStack` to `RasterStack`) and subsetting methods.

9. The discussion about memory footprint and the caching that `raster` does is relevant, but it would also be good to mention which dimensions a model can still have on e.g. a 4 Gb RAM machine. Programmers often forget how large and cheap RAM is, these days.

   We agree that this would be useful information, however, it is difficult to express the model dimensions in a meaningful way because it depends on several unknown factors, for example, the number of land use categories in the study region, and the number of explanatory factors. Instead, we have included the dimensions of the current example for PIE, since this is fairly typical of land use change modelling studies, and given the time to run the allocation procedure on a 4Gb RAM machine. We have added the following sentences to the third paragraph of the Discussion:

   "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."

10. P 10, L 22: "a raster object belonging to the raster package": rephrase.

    We have rephrased this sentence as follows:

    *"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 **raster** objects (RasterLayer, RasterStack or RasterBrick)."*

11. P 7 L 17: add to this sentence: "be expressed programmatically and be communicated as such with reasonable effort".

    We have made this addition. The full sentence now reads:

    "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)."

12. P 11, L 8,9: omit; put installation instructions in the Code availability section.

    We have done as the reviewer suggests. Also see response to comment 7.

13. Title: version should not be necessary. Suggest: *An open and extensible framework for spatially explicit land use change modelling: the* `lulcc` *R package*

    We have changed the title as suggested:

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

14. Under R 3.2.0, the source package passes R CMD check with only one (easy to resolve) NOTE

The NOTE has been resolved.

15. P 14, L 2: underscores are escaped; correct.

    We have fixed this in the revised manuscript.

16. P 15, L 12: avoid R comments, but use regular text.

    This has been corrected.

17. Why is Performance.plot not simply called plot?

    See response to comment 8.

18. How can the current script of the paper be run from within lulccR? Explain in "Code availability".

    The script of the paper is supplied as a demo with the package. We have added instructions about how to run the demo script to the Code availability section. In addition, we have added the following text to the first paragraph of Section 3:

    "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."

19. Are all figures reproduced by the scripts in this document? Or is there a demo script in lulcc that reproduces all figures in this paper?

    All plots can be reproduced by running a demo script included in the package: see response to the previous comment..

20. P 22, L 17,19: these expressions do not show a plot - why not call the method "plot", and show the plot instead of saving the plotting object here?

    We have attempted to manipulate the formatting restrictions as far as possible to make the Software description easier to follow. An alternative to the approach currently adopt would be to bundle the code with the figure, but in many cases this would remove the code from the description. Within the text we have attempted to point the reader to the correct Figure whenever "plot" is called: we hope this is a reasonable compromise.

Literature cited:

Aldwaik, S. Z. and Pontius, R. G. (2012). Intensity analysis to unify measurements of size and stationarity of land changes by interval, category, and transition. *Landscape and Urban Planning*, 106(1):103–114.

Beale, C. M., Lennon, J. J., Yearsley, J. M., Brewer, M. J., and Elston, D. A. (2010). Regression analysis of spatial data. *Ecology Letters*, 13(2):246–264.

Fuchs, R., Herold, M., Verburg, P. H., and Clevers, J. G. P. W. (2013). A high-resolution and harmonized model approach for reconstructing and analysing historic land changes in europe. *Biogeosciences*, 10(3):1543–1559.

Ince, D. C., Hatton, L., and Graham-Cumming, J. (2012). The case for open computer programs. *Nature*, 482(7386):485–488.

Mas, J., Kolb, M., Paegelow, M., Camacho Olmedo, M. T., and Houet, T. (2014). Inductive pattern-based land use/cover change models: A comparison of four software packages. *Environmental Modelling & Software*, 51:94–111.

Mascaro, J., Asner, G. P., Knapp, D. E., Kennedy-Bowdoin, T., Martin, R. E., Anderson, C., Higgins, M., and Chadwick, K. D. (2014). A tale of two "forests": Random forest machine learning aids tropical forest carbon mapping. *PLoS ONE*, 9(1):e85993.

MassGIS (2015). Massachusetts geographic information system, massgis.

Overmars, K., de Koning, G., and Veldkamp, A. (2003). Spatial autocorrelation in multi-scale land use models. *Ecological Modelling*, 164(2-3):257–270.

Pebesma, E., Nüst, D., and Bivand, R. (2012). The r software environment in reproducible geoscientific research. *Eos, Transactions American Geophysical Union*, 93(16):163–163.

Pontius, R. G. and Parmentier, B. (2014). Recommendations for using the relative operating characteristic (ROC). *Landscape Ecology*.

Pontius, R. G., Peethambaram, S., and Castella, J. (2011). Comparison of three maps at multiple resolutions: A case study of land change simulation in cho don district, vietnam. *Annals of the Association of American Geographers*, 101(1):45–62.

Pontius, R. G. and Schneider, L. C. (2001). Land-cover change model validation by an ROC method for the ipswich watershed, massachusetts, USA. *Agriculture, Ecosystems & Environment*, 85(1):239–248.

Rosa, I. M. D., Ahmed, S. E., and Ewers, R. M. (2014). The transparency, reliability and utility of tropical rainforest land-use and land-cover change models. *Global Change Biology*, 20(6):1707–1722.

Rosa, I. M. D., Purves, D., Souza, C., and Ewers, R. M. (2013). Predictive modelling of contagious deforestation in the brazilian amazon. *PLoS ONE*, 8(10):e77231.

Sing, T., Sander, O., Beerenwinkel, N., and Lengauer, T. (2005). ROCR: visualizing classifier performance in r. *Bioinformatics*, 21(20):3940–3941.

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

Veldkamp, A. and Fresco, L. (1996). Clue: a conceptual model to study the conversion of land use and its effects. *Ecological Modelling*, 85(2):253–270.

Verburg, P. H., De Koning, G. H. J., Kok, K., Veldkamp, A., and Bouma, J. (1999). A spatial explicit allocation procedure for modelling the pattern of land use change based upon actual land use. *Ecological modelling*, 116(1):45–61.

Verburg, P. H. and Overmars, K. P. (2009). 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 Ecology*, 24(9):1167–1181.

Verburg, P. H., Soepboer, W., Veldkamp, A., Limpiada, R., Espaldon, V., and Mastura, S. S. (2002). Modeling the spatial dynamics of regional land use: The CLUE-S model. *Environmental Management*, 30(3):391–405.

Wu, D., Liu, J., Zhang, G., Ding, W., Wang, W., and Wang, R. (2009). Incorporating spatial autocorrelation into cellular automata model: An application to the dynamics of chinese tamarisk (Tamarix chinensis lour.). *Ecological Modelling*, 220(24):3490–3498.

Manuscript prepared for Geosci. Model Dev. Discuss.
with version 2014/07/29 7.12 Copernicus papers of the LaTeX class copernicus.cls.
Date: 31 July 2015

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

~~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~~ In this paper we present the ~~lulccR~~ 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 paradigm: (1) ~~the~~ 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~~ Ensemble experiments to capture model structural uncertainty are difficult because of fundamental differences between implementations of ~~different~~ alternative models; (3) ~~different aspects of the modelling procedure must be performed in different environments because~~ Additional software is required because most 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~~ lulcc 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~~ 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 ~~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~~ (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 ~~model implementations~~ land use change modelling software 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). In this paper we describe the development of **lulcc**, a new R package designed to foster an open approach to land use change science.

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 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~~ modelling applications (Verburg et al., 2013). Under these circumstances it is useful to experiment with ~~different~~ various 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~~ is 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 PCRastersoftware~~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 such as map algebra and neighbourhood functions. Alternatively, the

PCRaster application programming interface (API) allows users to extend ~~the functionality of PCRaster in different~~ 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 model that simulates farmer decisions about agricultural land use in response to biophysical and socioeconomic driving factors~~deductive land use change model, 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~~ 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 different~~ , which handles several 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~~ implementations of CLUE-S ~~(Veldkamp and Fresco, 1996; Verburg et al., 1999), an earlier version of CLUE-S that operates at larger spatial scales,~~ and its predecessor, CLUE (Veldkamp and Fresco, 1996; Verburg et al., 1999), 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*~~ **sp** package (Pebesma and Bivand, 2005; Bivand et al., 2013) which unified many ~~different~~ alternative 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~~ **raster** package (Hijmans, 2014) provides many 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.~~ 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 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.

2 Design goals

The first design goal of ~~*lulccR*~~ **lulcc** is to provide a framework that allows users to perform ~~every stage~~ various stages of the modelling process ~~, shown by Fig. 1 ,~~ illustrated by Figure 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~~ seperate aspects of the procedure can easily be revisited. Thirdly, it ~~means it~~ is straightforward to ~~investigate the effect of different inputs and model setupson model outcomes~~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 ~~(Pebesma et al., 2012).~~ and be communicated as such with reasonable effort (Pebesma et al., 2012).

~~*lulccR*~~ **lulcc** is intended as an alternative to current paradigm of ~~closed-source~~closed source, specialised software ~~packages~~ 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 ~~data sets~~ 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 ~~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.~~ 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 ~~different~~ 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 ~~*lulccR*~~ **lulcc** 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~~ 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 ~~*lulccR*~~ **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 ~~diagram for *lulccR* together with a list of the most important functions~~ structure of **lulcc**, while Table 1 shows the functions included with the package. Here we describe the main components of ~~*lulccR*~~ **lulcc** integrated with an example application for the Plum Island Ecosystems dataset~~to demonstrate its functionality.~~ . 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 ~~*lulccR*~~ **lulcc** 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~~ The first of these contains data from 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 from the Boston,~~ , which in recent decades 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*~~ **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. 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 variablesare included~~1999 together with maps of three predictor variables: elevation, slope and distance to built land in 1985. ~~Land use for the site in 1985 is shown by Fig. **??**.~~

3.1.1 ~~Sibuyan~~

~~Sibuyan is a small island with a total area of 456 belonging to Romblon province~~ The second dataset includes information from Sibuyan Island 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~~ , and is a modified version of the dataset ~~distributed~~ supplied 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.~~ (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. ~~In *lulccR*~~ Currently **lulcc** requires all spatially explicit input data ~~must be stored in one of the file types supported by *rgdal* or exist~~ to exist either in the file system, in any of the formats supported by **raster**, or in the R workspace as ~~a raster object belonging to the *raster* package~~ **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 ~~typically~~ usually 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~~ time point is required, as shown by ~~Fig.~~Figure 1. Ideally, two more observed land use maps for subsequent ~~timesteps~~ 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.

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

```
> 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 ~~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~~ time points associated with the observed land use maps. The first ~~timestep~~ time point must always be zero; if additional maps are present they should ~~have timesteps~~ be associated with time points greater than zero, even in backcast models. In most land use change modelling applications ~~a timestep~~ 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 ~~two~~ observed land use maps from ~~different times~~ 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 ~~*lulccR*~~ **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
```

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

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

Apart from observed land use ~~maps and predictor~~ and explanatory variables other input maps may be required. The two allocation routines currently included with ~~*lulccR*~~ **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 ~~*lulccR*~~ **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 ~~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,~~ 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 ~~to inductive~~ (e.g. Tayyebi et al., 2014). One reason why R is attractive for land use change modelling ~~(e.g. Tayyebi et al., 2014). Currently~~ ~~*lulccR*~~ 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*~~ **rpart** package (Therneau et al., 2014), and random forests, provided by the ~~*randomForest*~~ **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~~ , assume the error terms of 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~~ There is also some evidence that 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~~ may 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~~ Mascaro et al. (2014), even though they do not assume independence. A simple approach to reduce the impact of this phenomenon is to fit ~~the~~ predictive 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~~ In the following code snippet we create training and testing partitions for the Plum Island Ecosystems dataset by performing a stratified random sample~~of the data . The data partition is obtained as follows:~~. 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,
                Forest~ef_001+ef_002,
                Other~ef_001+ef_002)

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

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

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

~~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 datastructure 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~~ The model fitting functions each return an object of class PredictiveModelList containing a predictive model for each land use ~~. A typical workflow is shown here: 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~~ type. With these objects it straightforward to map the suitability of 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~~ over the study region. 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)
> points   <- rasterToPoints(obs[[1]], spatial=TRUE)
> probmaps <- SpatialPointsDataFrame(points, probmaps)
> probmaps <- rasterize(x=probmaps, y=obs[[1]],
                        field=names(probmaps))
```

```
> levelplot(probmaps)
```

In some circumstances it may be appropriate to supply 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~~ 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 ~~permits a null model to be fitted. Predictive models for each land use are represented by an object of class `PredModels`:The resulting object makes it straightforward to plot the suitability of each land use over the study region using the~~ model with no explanatory variables. To do this, a formula such as `Forest~1` should be supplied to the ~~`calcProb`~~`glmModels` 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*~~ **ROCR** package (Sing et al., 2005), allowing the user to assess model performance using ~~several~~ various 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.~~ fit.

In ~~*lulccR*~~ **lulcc** we extend the native ~~*ROCR*~~ **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~~ prediction and performance classes of **ROCR** are extended by PredictionList and PerformanceList, respectively, to handle objects of class PredictiveModelList. The procedure to evaluate several ~~`PredModels` objects~~ 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)
> rpart.perf <- PerformanceList(pred=rpart.pred,
                                measure="rch")
> rf.pref    <- PredictionList(models=rf.models,
                               newdata=test.data)
> rf.perf    <- PerformanceList(pred=rf.pred,
                                measure="rch")
> plot(list(glm=glm.perf, rpart=rpart.perf, rf=rf.perf))
```

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

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 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,
                         spatial=TRUE)
> test.data <- getPredictiveModelInputData(obs=obs,
                                           ef=ef,
                                           cells=part,
                                           t=6)
> glm.pred  <- Prediction(models=glm.models[[2]],
                          newdata=test.data)
> glm.perf  <- Performance(pred=glm.pred,
                           measure="rch")
> 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 ~~the total area~~ either the total number of cells occupied by each ~~land use type at every~~

~~timestep~~category or the number of transitions among the various categories at each time point. 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~~ time point based on the spatial configuration of the landscape at the previous ~~timestep~~ time point (e.g. Rosa et al., 2013), it is more common to ~~supply land use area for every timestep~~ 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). ~~This is the approach currently supported in *lulccR*.~~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 data.frame before running the allocation routine.

Land use area may be estimated using non-spatial land use models or, ~~if the study aims to reconstruct historic land use change~~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). ~~*lulccR*~~ **lulcc** 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~~ time points for which observed land use data is available. However, doing so is useful for model ~~validation: we~~ pattern validation, allowing us to test the ability of ~~the model~~ models to predict the ~~location~~ spatial allocation of change given the exact quantity of change.

3.5 Allocation

~~The allocation component of~~ The allocation algorithm in land use change models ~~estimates the location of change in the study region at each timestep~~ determines the pixels in which various land use transitions should take place (Verburg et al., 2002). Currently ~~*lulccR* includes two inductive~~ **lulcc** includes two 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. 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*.~~ 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.

3.5.1 Decision rules

The first decision rule included in ~~*lulccR*~~ **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 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,~~ , such as 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~~(Verburg and Overmars, 2009). 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~~ that 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~~ three decision rules are ~~implemented~~ applied by the `allow` function ~~while the fifth decision rule is performed~~ 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 ~~*lulccR*~~ **lulcc** these are represented by the ~~`NeighbMaps`~~`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 ~~timestep.~~ time point.

3.5.2 CLUE-S allocation method

The CLUE-S model implements an iterative procedure to meet the specified demand at each ~~timestep~~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.

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*~~ **lulcc** allows any predictive model supported by ~~`PredModels`~~ PredictiveModelList to be used. ~~After this step the suitability is increased for land uses where the~~ For each land use the algorithm determines whether the allocated area is less than~~demand and decreased for land uses where it is~~, equal to or greater than the specified demand. If it is less than or greater than demand ~~. The extent to which the suitability~~ the suitability of each pixel in the study region to the land use in question is increased or decreased~~is a function of~~, respectively, by an amount depending on the difference between ~~allocated change and demand.~~ 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 uses, within a user-defined tolerance, is met. ~~The~~ At each iteration the original model perturbs the ~~location suitability to~~ 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 ~~*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~~ **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 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 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`~~ 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,
                      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,
                            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. **??**, 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~~ algorithm because it doesn't simulate competition between ~~different land use types~~land uses. 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~~ 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. Here, n cells with the lowest ~~location~~ 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. ~~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~~ 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 ~~location~~ 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 ~~*lulccR* the ordered~~ **lulcc** the ordered allocation model is represented by the ~~`OrderedModel`~~ OrderedModel class. In the following code we create an ~~`OrderedModel`~~ OrderedModel object, supplying the order in which to allocate change (built, forest, other), and pass this to the ~~generic~~ `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 ~~Validation~~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 ~~timestep~~ time point (Pontius et al., 2011). Previous studies have extracted useful information from the three possible two-map comparisons ~~(e.g. **?**)~~(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 simulataneously. 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 ~~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~~ useful to perform pattern validation 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*~~**lulcc**, three-dimensional contingency tables at ~~different~~ multiple resolutions are represented by the ~~`ThreeMapComparison`~~ ThreeMapComparison class. Two subclasses of ~~`ThreeMapComparison` represent different~~ ThreeMapComparison represent two types of information that can be extracted from the tables: ~~the `AgreementBudget` class~~ AgreementBudget represents sources of agreement and disagreement between the three maps at ~~different resolutions while the `FigureOfMerit` class~~ several resolutions 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 ~~`AgreementBudget` and `FigureOfMerit`~~ ThreeMapComparison, 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,
                                     factors=2^(1:8),
                                     timestep=14)
> 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)
```

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

4 Discussion

The example application for Plum Island Ecosystems demonstrates the key strengths of the ~~*lulccR*~~ **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 ~~packages~~programs. 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~~ various 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*~~ **lulcc** is substantially different from alternative environmental modelling frameworks. Most ~~significantly, *lulccR*~~ 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 ~~different~~ various 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~~ most software programs for land use change ~~models~~ modelling are designed with the user in mind, with very few ~~models~~ providing any way for users or developers to improve or even understand ~~the model implementation. With *lulccR*~~ 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 ~~*lulccR*~~ **lulcc** will work across Windows, ~~Mac OS X~~ MacOS and Unix platforms, whereas many existing applications are platform dependent. Comprehensive documentation of the functions, classes and methods of ~~*lulccR*~~**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.

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*~~ **raster** package overcomes this limitation by storing raster files on disk and processing data in chunks ~~(Hijmans, 2014). *lulccR*~~ (Hijmans, 2014). **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 provides~~ **raster** package offers one possible solution to this problem, allowing ~~users to make predictions with predictive models~~ predictive models to be used in a memory-safe way. In effect, this would mean spatially explicit input data including observed land use maps and ~~predictor~~ 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 ~~*lulccR*~~ **lulcc** 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*~~ **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 ~~timestep~~ 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 ~~location~~ 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 ~~location~~ allocation of change simultaneously, whereas the current routines in ~~*lulccR*~~ **lulcc** only consider the ~~location~~ 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 ~~*lulccR*~~ **lulcc** with this class of model would be a valuable addition to the software because land use change is increasingly recognised as ~~a regionaland global issue that occurs over multiple scales.~~ 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 interpretation of model diagnostic ability.

One of the main strengths of ~~*lulccR*~~ **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 ~~*lulccR*~~ **lulcc** to broaden the ~~different~~ available model structures and, therefore, improve the ability of ~~*lulccR*~~ **lulcc** to capture model structural uncertainty. The methods in the current version of ~~*lulccR*~~ **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). ~~The main advantage of these models is thatthey are able~~ This means that, unlike inductive models, they can be used to establish causality ~~because they allow modellers to test specific theories about the location of change and predictor variables whereas inductive models simply associate~~ between 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~~ and its driving factors (Overmars et al., 2007). Including this class of model ~~would permit multiscale studies whereby~~ in **lulcc** would allow inductive and deductive land use change models ~~operating at~~ with different spatial resolutions ~~are~~ 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 ~~encourages~~ 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 ~~*lulccR*~~**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). ~~Of course, users~~ Users may also develop their own R packages that depend on ~~*lulccR*~~ **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 ~~predictor~~ explanatory 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*~~**lulcc**, a free and open source software package providing an object-oriented framework for land use change modelling in R. ~~*lulccR* allows the entire~~ **lulcc** allows various aspects of the modelling process to be performed within the same environment, supports three ~~different~~ 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 ~~are able to~~ can 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*~~ **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 ~~*lulccR*~~ **lulcc** source code currently resides on ~~GitHub: .~~ 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")
```

The Supplement related to this article is available online at doi:10.5194/gmdd-0-1-2015-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.

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.

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.

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.

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 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., 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 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.

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.

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.

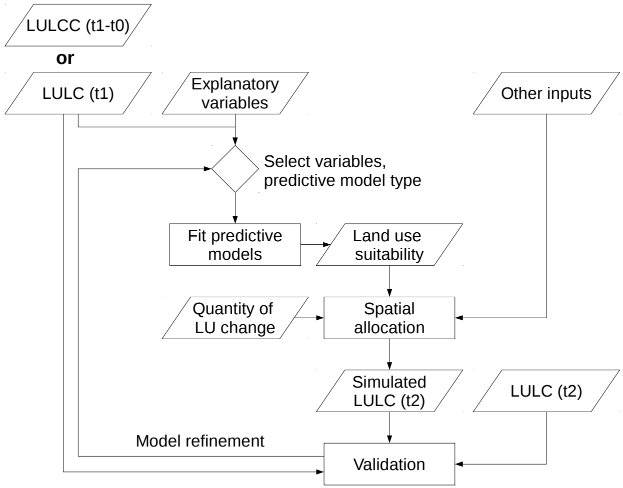


Figure 1. Diagram showing the general methodology used for inductive land use change modelling applications, adapted from Mas et al. (2014). The input land use/land cover data can be a single categorical map showing the pattern of land use/land cover at one time point (LULC (t1)) or a series of maps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

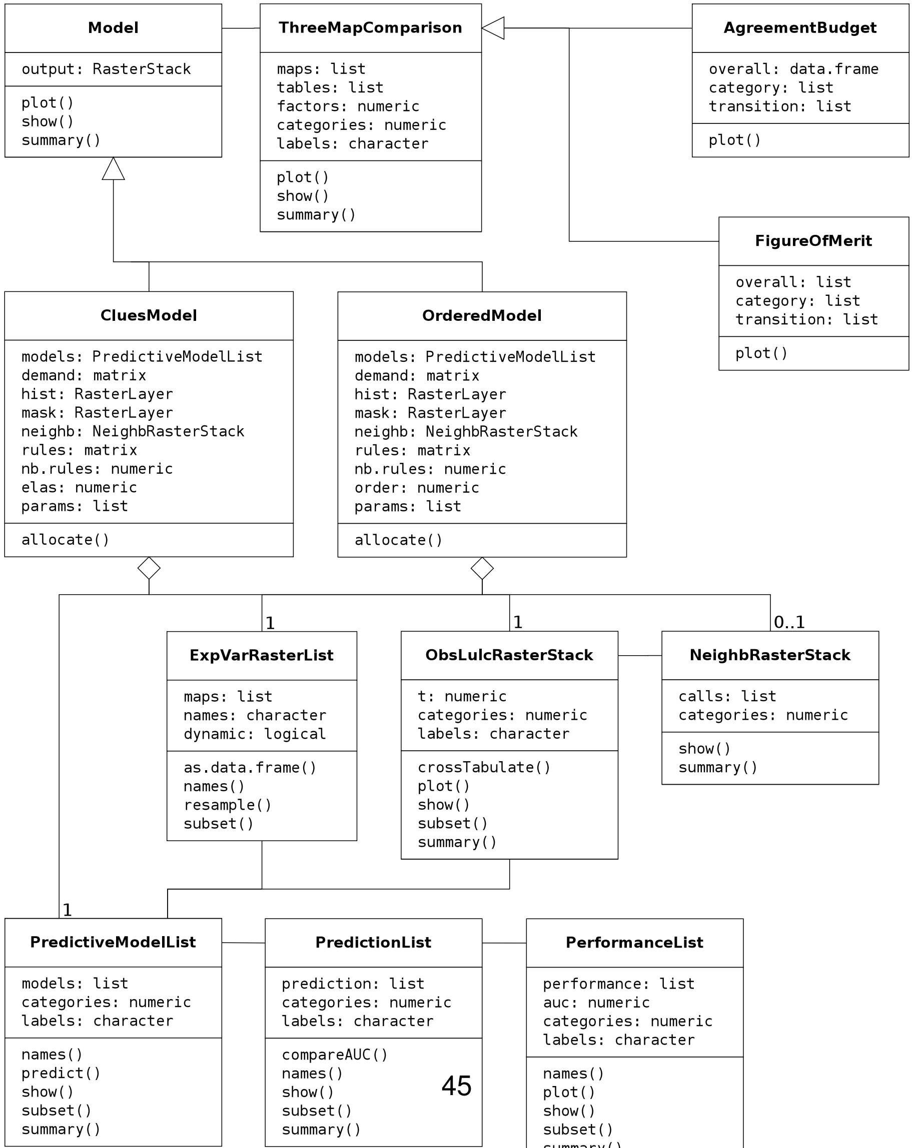

Model
output: RasterStack
plot()
show()
summary()
ThreeMapComparison
maps: list
tables: list
factors: numeric
categories: numeric
labels: character
plot()
show()
summary()
AgreementBudget
overall: data.frame
category: list
transition: list
plot()
FigureOfMerit
overall: list
category: list
transition: 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
1
0..1
ExpVarRasterList
maps: list
names: character
dynamic: logical
as.data.frame()
names()
resample()
subset()
ObsLulcRasterStack
t: numeric
categories: numeric
labels: character
crossTabulate()
plot()
show()
subset()
summary()
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()

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

Table 1. ~~Plum Island Ecosystems site land use map for 1985. In recent years~~ Functions included in the ~~site has undergone extensive land use change from forest to built areas.~~ **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

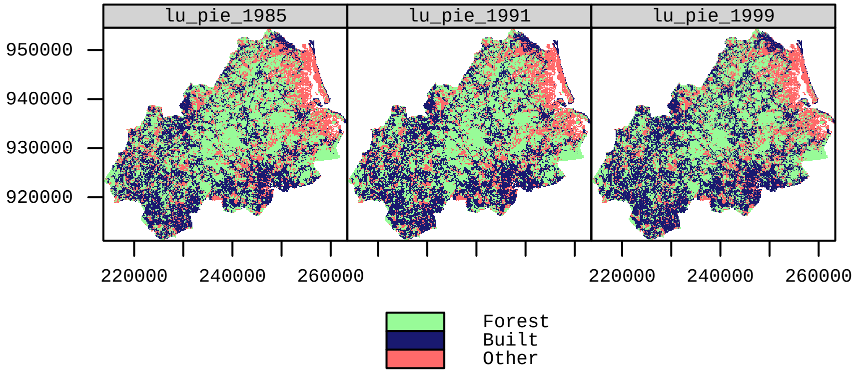


Figure 3. Observed land use maps for the Plum Island Ecosystems site in 1985, 1991 and 1999, created by plotting the ObsLulcRasterStack object representing the data.

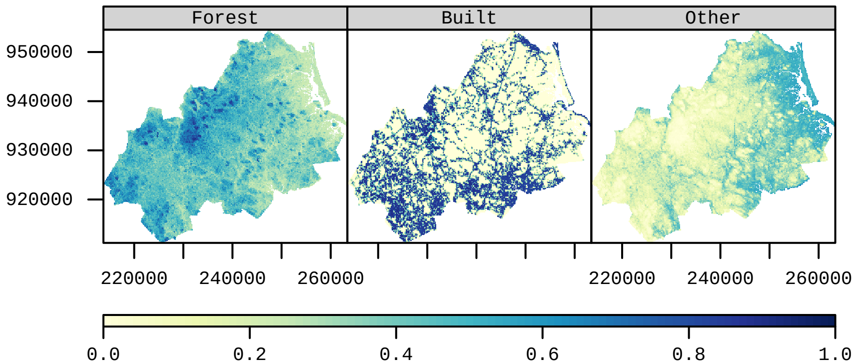


Figure 4. ~~Land use suitability maps for~~ Suitability of pixels in the Plum ~~Islands~~ Island Ecosystems study ~~area. The “forest”~~ site to Forest, Built and Other land use ~~class has uniform suitability because we employ a null model. Occurrence of “built” is related~~ classes according to ~~elevation, slope~~ binary logistic regression models. Elevation and slope are used as explanatory variables for all land uses while Built additionally includes distance to ~~1971~~ built ~~area, while “other” is related to elevation and slope only.~~ pixels in 1985.

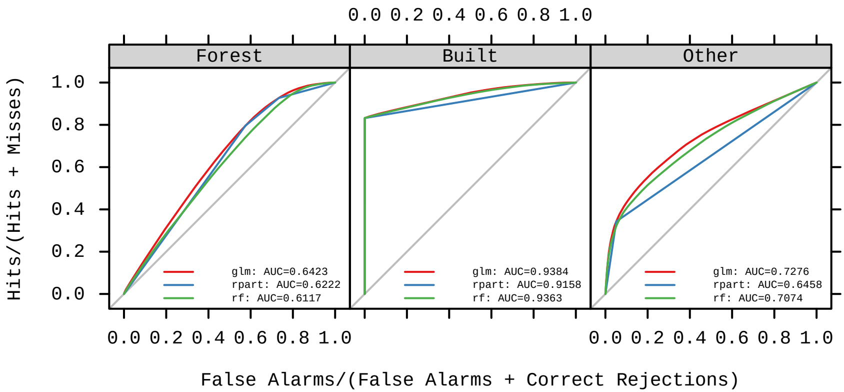


Figure 5. ROC curves ~~for~~ showing the ability of each ~~statistical~~ type of predictive model ~~for each~~ to simulate the observed pattern of land use ~~. Note that because~~ in the ~~forest class employs a null model only~~ Plum Island Ecosystems site in 1985 in the ~~logistic regression model is calculated~~data partition left out of the fitting procedure.

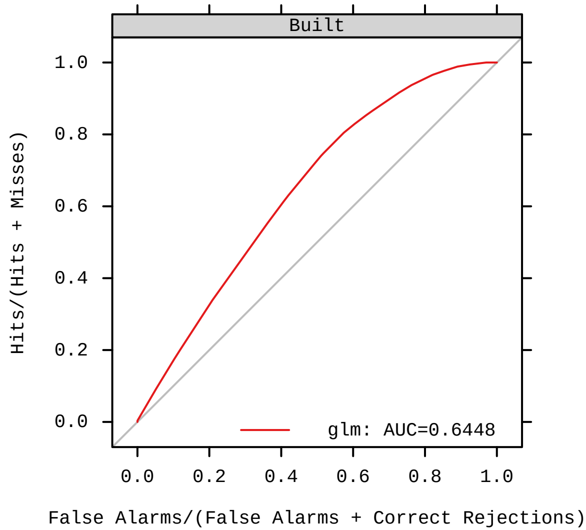
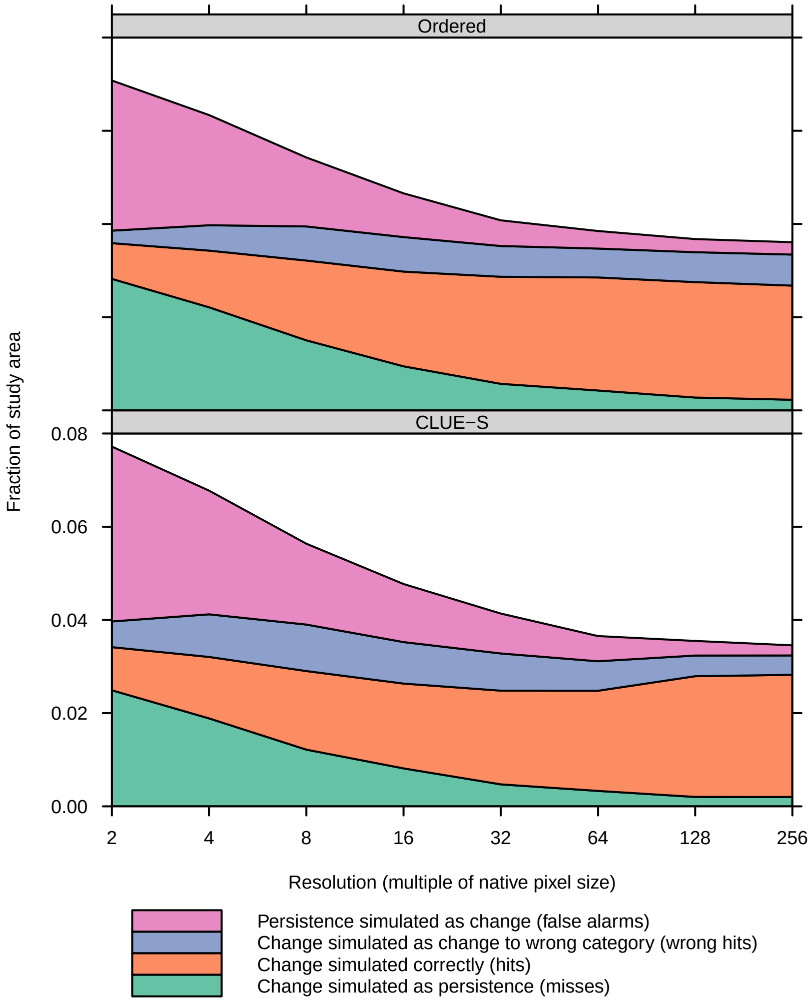


Figure 6. ~~Overall agreement budget comparing~~ ROC curve showing the ~~*lulccR* CLUE-S algorithm with~~ ability of the ~~original~~ binary logistic regression model ~~output for 2011. This shows a good level of agreement between~~ fitted on observed land use data from 1985 to predict the ~~two maps: the proportion of persistence~~ gain in Built land between 1985 and ~~change simulated correctly is high compared to incorrectly simulated persistence or change.~~1991.



Ordered
CLUE−S
0.08
0.06
0.04
0.02
0.00
Fraction of study area
2
4
8
16
32
64
128
256
Resolution (multiple of native pixel size)
Persistence simulated as change (false alarms)
Change simulated as change to wrong category (wrong hits)
Change simulated correctly (hits)
Change simulated as persistence (misses)

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.

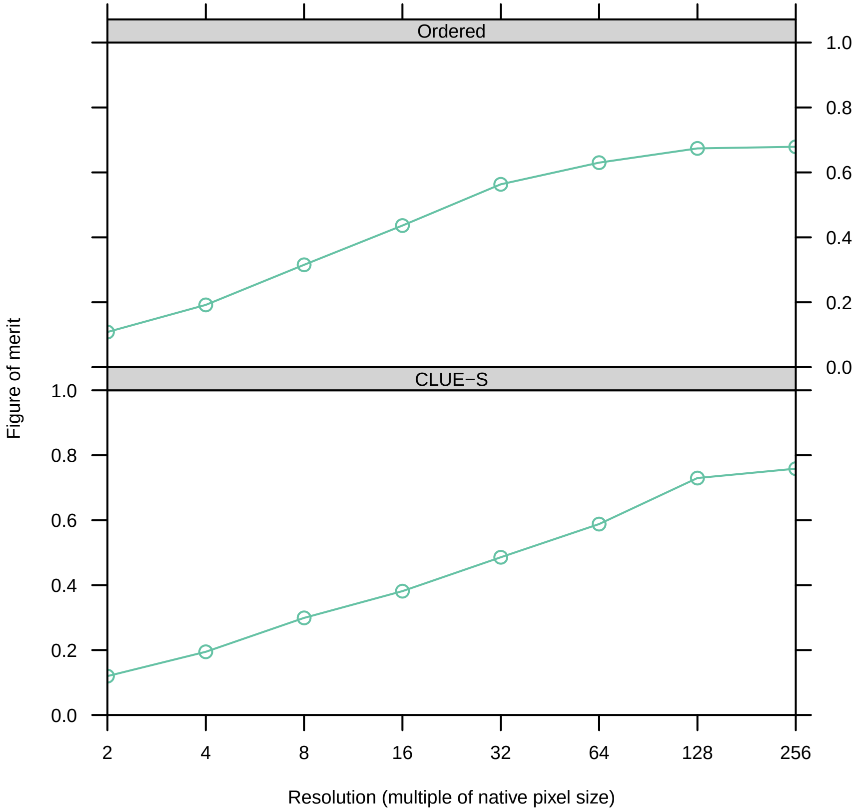


Figure 8. Figure of merit scores corresponding to the agreement budgets depicted in ~~Fig.~~Figure 7.