Review for:	Geoscientific Model Development
Date:	16 May 2015
Manuscript number:	gmdd-8-3359-2015
Manuscript Title:	An open and extensible framework for spatially explicit land use change modelling in R: the lulccR package (0.1.0)

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 this 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 target audience. This manuscript does not need to describe why land change is important. Moreover, the manuscript does not even need to describe why modeling is important. The manuscript must focus on describing why land change modelers might want to use the authors' software.

The remainder of this review is in order of the sections of the manuscript.

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?"

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

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.

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.

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.

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

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 distracted. 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 example application should suffice. Two case studies would 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.

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.

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 meaning of “correct spatial resolution” is not clear on page 3369.

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 …”

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

In line 5 of page 3372, should “occurrence” be “gain”?

It is extremely confusing to use 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 modeling. 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 me 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.

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.

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 Si (2014). My students have created R code for TOC.

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.

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

Section 3.5.1 must discuss how the algorithms deals 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 Other, then how does the software decide whether Built or Other gains from the Forest pixel?

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

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 “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.

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. 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 that 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.

The title of Section 3.6 should be "Pattern Validation" rather than "Validation", to distinguish from Process Validation. In section 3.6, "Pontius et al. 2007" should be "Pontius et al. 2008". Line 14 of page 3379 should change from "allocation performance" to "quantity and allocation performance". 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.

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 one constraint to the advancement of the science of land change modeling.

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

Figure 4 must say the suitability for what?

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 candidates 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 those pixels must be eliminated from the ROC analysis. The shape of the curve for Built in figure 3 makes be believe that the authors did not eliminate those pixels. This is a common blunder in the profession. Figure 3 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.

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.

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 the full range. I do not see any need for numbers on the right axis.

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.

LITERATURE available at www.clarku.edu/~rpontius

Pontius Jr, Robert Gilmore and Marco Millones. 2011. Death to Kappa: birth of quantity disagreement and allocation disagreement for accuracy assessment. International Journal of Remote Sensing 32(15): 4407-4429.

Pontius Jr, Robert Gilmore and Kangping Si. 2014. The Total Operating Characteristic to measure diagnostic ability for multiple thresholds. International Journal of Geographical Information Science 28(3): 570-583.

Pontius Jr, Robert Gilmore and Benoit Parmentier. 2014. Recommendations for using the Relative Operating Characteristic (ROC). Landscape Ecology 29(3): 367-382.