Biogeosciences Discuss., 7, C4368–C4385, 2010
www.biogeosciences-discuss.net/7/C4368/2010/

Biogeosciences Discussions

Interactive comment on "Integrating field sampling, spatial statistics and remote sensing to map wetland vegetation in the Pantanal, Brazil" *by* J. Arieira et al.

J. Arieira et al.

juarieira@ufmt.br

Received and published: 16 December 2010

Response to the reviewer

Referee 2#

Authors: We would like to thank the referee for his/her review and the useful suggestions.

General comments

R#2: The study shows an interesting combination of advanced methods for mapping wetland vegetation which can help to map vegetation over large areas to generate

Full Screen / Esc

Printer-friendly Version

Interactive Discussion

Discussion Paper

relevant input data for dynamic models, especially when field mapping is restricted by available resources or by areas that are difficult to access. Points of improvement are the structure of the paper, reference to other remote sensing studies and the potential role of the mapping procedure for wetland conservation. It is a valuable contribution in the field of bio-geosciences and should therefore be considered for publication if a number of changes are applied as specified below.

Authors: The reviewer recognizes the relevance of our work described in the paper. We generally agree with his/her suggestions for improvement. Below we provide a discussion on the specific comments providing details on what has been changed in the revised manuscript.

Specific comments

Structure of the paper

R#2: I would recommend a few changes in the structure of the paper, because now data section and analysis section are entangled; I would move the cluster analysis to the section 6, as this is also a part of the mapping and move parts of section 6 to the data section, e.g. which kind of remote sensing are used, description of the DEM etc.

Authors: In the revised manuscript, we moved cluster analysis to section 6.5 (Spatial distribution of plant communities across the floodplain) and part of the section 6, specifically the acquisition of image data, to the data section 4 (Data acquisition).

Validation of mapping results

R#2: The authors applied 2 different methods to evaluate the quality of the mapping result. Is it important to include both approaches or would 1 be sufficient. And if yes, which one is preferred and why?

Authors: We think that both methods for evaluating the quality of the statistical approach for mapping are important. While cross-validation evaluates classification accuracy by comparing mapped and observed vegetation communities at sampling lo-

BGD

7, C4368–C4385, 2010

Interactive Comment

cations, Monte Carlo simulation evaluates the uncertainty in the mapping procedure itself, by verifying the stability of predictions using realizations of maps instead of mean Kriged maps. Also, the cross-validation evaluates the effect of the configuration and density of observations on mapping accuracy, which is not done in the Monte Carlo simulation.

Discussion

R#2: I would start this section with the research questions and then use the analysis results to answer the research questions. The first sentences of the discussion could be the start of the conclusion section which is currently incorporated in the discussion. In the conclusion section the authors should repeat the main findings and suggestions for future studies.

Authors: We created a conclusion section to provide the main findings of our study and provide suggestions for future studies; we started the discussion section highlighting the research questions.

R#2: Cluster analysis is based on 7 classes, however, no information is provided for the choice of 7. We added this information to the revised manuscript.

Authors: We provided information on how the seven classes are defined as follows. The seven clusters were determined based on two factors: 1) the proximity of point scores on the ordination space, 2) the general expert knowledge (including existing vegetation community classifications for the area cited in the manuscript) of vegetation communities in the area.

R#2: Flooding duration is an important driver of vegetation zonation – would you recommend to include that factor in the prediction?

Authors: The inclusion of flooding duration as a predictor variable in the model depends on the objectives of the research. If the flooding duration is included as predictor, this information can not be used in analysis of the relation between flooding pattern and vegetation community pattern (as is done here), because the information of flooding pattern would be contained within the vegetation map. Our interest was to compare information of inundation and vegetation distribution as in gradient analysis but using continuous data provided by both the inundation and the vegetation maps, instead of using observation points only. Nevertheless, because inundation is an important driver of plant community distribution, including flooding duration in the model may improve map quality. The gain in quality of the map would however be limited because flooding duration is only known with a relatively high uncertainty (as it depends on complicated flow dynamics on the floodplain). Nevertheless, we shortly discussed (in the Discussion section of the revised manuscript) the possibility of using flooding duration in the mapping procedure.

R#2: Remote sensing data where used as input for universal kriging; what is the advantage /disadvantage of this approach to supervised classification?

Authors: Supervised classification is the general name for classification methods in remote sensing where the first step is to indicate pixels with known classes and the second step to assign the other pixels to one of those classes based on their (spectral) similarity. Many different methods exist, the maximum likelihood being used most often. These classification methods have a very limited capacity to take spatial dependence variance into account. Universal kriging is an interpolation method integrating observations from different data sources and taking into account the spatial autocorrelation observed. The vegetation in our study area shows complex spatial patterns with significant autocorrelation while spatially distributed observations are available that can be used for interpolation, hence universal kriging is the natural choice. This was added to the Discussion section.

R#2: What is the relation of this particular research to wetland conservation?

Authors: The value of this research for wetlands conservation is twofold: 1) the study area has a large ecological value, and 2) the mapping procedures used are relevant for wetland conservation in general. The ecological, social and economic importance of wetlands was internationally recognized in the Ramsar Convention. Around 20% of the Brazilian territory is covered with wetlands, but so far there is no police guidance regulating the sustainable use of these wetlands. As a result, wetlands of overwhelming ecological importance, as is the case of the Pantanal Matogrossense studied here, remain threatened by human activities. Species diversity of the Pantanal is associated with large habitat diversity indicating that habitat conservation is a topic of high conservation priority (Junk et al. 2006). Thus, the conservation of the Pantanal needs information on the type and distribution of their habitats, required for monitoring impacts and defining conservation strategies. Wetland mapping at scales suitable for conservation can be limited by the lack of temporally and spatially consistent dataset. The present study aims at contributing to identify efficient methodological approaches to identify and describe spatio-temporal patterns of wetland vegetation of the Pantanal. We provide a framework to map wetland habitats based on the integration of field data and remote sensing data through geostatistical methods and identify possible causes of vegetation zonation. These two items were highlighted in the introduction.

R#2: Given this approach is used for monitoring, do we need to re-do the field sampling? R#2: How to go about with only RS images and the DEM?

Authors: The spatial predictions of vegetation are based on the capability of remote sensing images and ancillary data to describe vegetation patterns observed in stands. We assume in this work that plant communities arise from the combination of patterns of distribution of different life forms and that these patterns repeat themselves in other parts of the Pantanal and present a temporal stability. Because the vegetation patterns captured in our field sampling should repeat themselves in time and in other areas, one would be able to monitor vegetation patterns by just providing image data (RS images and the DEM) to create vegetation maps for different years. However, the assumption of temporal stability in vegetation patterns is constrained by long-term dynamics associated with events (e.g. gradual increase in flooding durations due to climate change

BGD

7, C4368–C4385, 2010

or land use change in upstream areas) that can change plant associations and interactions. To deal with this, we would suggest that field sampling is re-done at a regular time interval of approximately ten years. We included this in Discussion section (revised manuscript).

Technical

6890

Abstract

R#2: First sentence needs re-writing, because abstract starts with reference to wetland protection, but the core of the study is the development of a methodology for mapping wetland vegetation and therefore the abstract should start with the main message concerning the larger part of the paper; and furthermore, the first sentence contains 2 aims.

Authors: This was changed in the revised manuscript.

R#2: 4: change sophisticated to advanced; interpolation and error propagation are also statistical techniques; authors could summarize this as 'advanced statistical techniques', in particular.

Authors: This was changed in the revised manuscript.

R#2: 5: change to: . . . .to describe spatial vegetation patterns

Authors: This was changed in the revised manuscript.

R#2: 19: change to: . . . .derived from Monte Carlo

Authors: This was changed in the revised manuscript.

R#2: 25: what are new digital images?

Authors: This was changed in the revised manuscript to: . . .by including a larger number of digital images as explanatory variables in the model.

BGD

7, C4368–C4385, 2010

Interactive Comment

6891

Abstract

R#2: 1: providing a new basis

Authors: This was changed in the revised manuscript.

R#2: 2: allow – remove 'ing'

Authors: This was changed in the revised manuscript.

Introduction

R#2: 5: suggestion: wetlands are vulnerable habitats, threatened by. . .

Authors: This was changed in the revised manuscript.

R#2: 12: Is it indeed lack of knowledge?

Authors: This was changed in the revised manuscript to:. . .lack of police guidance to regulate the sustainable use of wetlands may lead. . .

R#2: 18: Two aspects (because factors are also used for something else)

Authors: This was changed in the revised manuscript.

6892

Introduction

R#2: 3: specify the environmental gradients

Authors: This was changed in the revised manuscript to: . . .environmental gradients (e.g. altitude, temperature, water, nutrients) . . ..

R#2: 11: remote sensing analysis enables

Authors: This was changed in the revised manuscript.

BGD

7, C4368–C4385, 2010

Interactive Comment

R#2: 19: Point data or other field data

Authors: This was changed in the revised manuscript.

6893

Introduction

R#2: 1: If I understand correctly, emphasis is actually on understanding the complexity, and not only the effect of flooding.

Authors: This was changed in the revised manuscript to: . . . to understand the complexity of spatial patterns of vegetation distribution and to study . . .

R#2: 11: What are new data? Probably you mean field-data

R#2: 10: Aim 1 and 2 could be combined; . . . to identify plant communities by combining field data and remote sensing data using advanced statistical techniques

Authors: This was changed in the revised manuscript to: The aims of this paper are: 1) to identify and map plant communities of the Pantanal by combining field data and remotely sensed data using advanced statistical techniques; 2). . .

R#2: 14: leave out 'on the basis of'

Authors: This was changed in the revised manuscript.

6893

Study area

R#2: 18: start with the location of the study area

Authors: This was changed in the revised manuscript.

R#2: 26: trend in precipitation – omit 'the'

Authors: This was changed in the revised manuscript.

BGD

7, C4368–C4385, 2010

Interactive Comment

R#2: 27: change topography to ‘elevation’

Authors: This was changed in the revised manuscript.

6894

Study area

R#2: 5: i.e. (remove’,’)

Authors: This was changed in the revised manuscript.

R#2: 19: change ‘over’ to ‘of’

Authors: This was changed in the revised manuscript.

Outline of the approach

R#2: I would rephrase this part: The first step was. . . .The second step. . .

Authors: This was changed in the revised manuscript.

R#2: 24: I would avoid high-resolution field sampling as high-resolution is often used for RS images;

Authors: This was changed in the revised manuscript by removing high-resolution.

R#2: 27: Start with: Remote. . . .; omit ‘these’; change ‘are’ to ‘were’ if you want to say what has been done;

Authors: This was changed in the revised manuscript.

6895

Outline of the approach

R#2: 2: Change ‘these regressions’ into ‘the regression analysis’

Authors: This was changed in the revised manuscript.

BGD
7, C4368–C4385, 2010

R#2: 3: was performed; . . . to combine extracted factor scores and spatially continuous information derived from

Authors: This was changed in the revised manuscript.

Field data

R#2: 12: We have sampled vegetation data based etc.; there is no need to talk about classification, as this section is about the data; the classification should go into the analysis section;

Authors: This was changed in the revised manuscript.

R#2: 23: focus on; remove 1 'on'

Authors: This was changed in the revised manuscript.

6896

Field data

R#2: 6: Specify more the sampling scheme

Authors: This was changed in the revised manuscript by providing more detail information on the sampling scheme according with the answers below.

R#2: 8: 23 trails each of 250m

Authors: This was changed in the revised manuscript.

R#2: 13: why 50m?

Authors: Because we wanted to sub-divide each trail in regular intervals and that these sub-divisions presented a length suitable for sampling large trees. This information was provided in the revised manuscript.

R#2: 15: what is measurement acquisition?

BGD

7, C4368–C4385, 2010

Interactive Comment

Authors: This was changed to data collection

R#2: 17: do you mean point quadrat or points along the trail?

Authors: Bullock (1996) called the procedure of counting of plant individuals that touched on a vertical rod (or pin), established in a determined distance interval along transects by point quadrat method. This method provides precise estimates of cover.

R#2: 19-21: what do you mean with ‘intercepted by the plant’?

Authors: We mean that all plant individuals (herbs and vines) that touched on the vertical rod established in 2 m intervals along each trail were included in the sampling.

R#2: Section 4.2: I would include a table related to the sampling scheme to get an overview in terms of plot size, frequency etc.

Authors: We included a table related to the sampling scheme

6897

R#2: 19: How does such a quantification look like?

Authors: We left quantification out and changed to:. . .by means of structure and composition data (only dominant species) of different vegetation layers.

R#2: 21: Defining plant communities: where are the clusters coming from and why 7 clusters?

Authors: The seven clusters were determined based on two factors: 1) the proximity of point scores on the ordination space, 2) the general knowledge of vegetation communities in the area.

6999

Mapping plant communities

R#2: 23: would change ‘accurate’ into ‘detailed’

BGD

7, C4368–C4385, 2010

Interactive Comment

Authors: This was changed in the revised manuscript.

6900

Mapping plant communities

R#2: 13: It is unclear to me why both PCA factors as well as original bands are used; in my understanding you would derive PCA factors to reduce the number of inputs and to derive meaningful factors with which you would use for further analysis; so why using both? And what is the meaning of each PCA factor?

Authors: We expected that this would occur, that PCA factors would provide meaningful factors and replace band data in the model. This would occur because PCA factors summarize spectral information and are correlated with the bands. PC1 expresses a gradient of brightness; PC2 a gradient of greenness; and PC3 of wetness. For this reason, we would expect that the models were composed by either spectral bands or PCA factors. However because our more direct interest in using multiple regression analysis in this study was to predict values, and not properly to test hypotheses about the model parameters, collinearity of explanatory variables was not a matter of great concern (Legendre and Legendre 1998). For this reason, we decided using an automatic variable selection technique to choose regression models that provided the best combination of parameters with the strongest contribution for prediction. In our case, the best models included both Principal Component factors and bands. We have tested the uncertainty on the estimations using Monte Carlo simulation.

R#2: 17: reference to the other study by Ozesmi and Bauer should be made in the literature review or discussion; I don't consider it important here.

Authors: This was changed in the revised manuscript.

R#2: 19: SRTM data: is it a problem that the DEM measures the top of the trees? – please comment on that in the discussion section

Authors: Our main interest in using Remote sensing imagery and SRTM DEM was to

BGD

7, C4368–C4385, 2010

Interactive Comment

include them as explanatory variables to improve the model ability to make accurate spatial predictions on vegetation patterns. Because testing hypothesis on the relationship between environmental variables and vegetation patterns was not the objective of this study, it was not a problem that DEM provided measurement of the top of the trees. This is discussed in the revised manuscript.

R#2: 24: would refer to pixel size when referring to the RS image; how to go about the different cell-sizes? Is it a problem to disaggregate a 90m cell-size to 40x50m? – to be addressed in discussion section.

Authors: The disaggregation of the 90m cells into 40x50m cells was done by overlaying the desired 40x50m cells over the 90m cells and creating intersects. From these a weighted average was calculated to determine the value for the 40x50m cells. In this study, instead of assuming that land cover fits very well into multiple pixels and such areas are homogeneous up to the coarsest pixel (SRTM DEM) (Fisher 1997), we have assumed that spatial structure of data affects classification accuracy and used a geostatistic model that accounts the spatial structure using variograms of the modeled error. Image variance determined the ability to extract useful information about scene conditions (Collins and Woodcock 1999).

6901

Mapping plant communities

R#2: 3-4: What is an interactive raster GIS environment compared to other raster GIS software?

Authors: The difference between an interactive raster GIS environment and raster GIS software is that the first one is able to create interactive spatio-temporal environmental models, while the second is focused on map algebra and store spatial data. This was commented in the revised manuscript.

R#2: 8: What are the DEM-derived variables? And why Pearson? Is the relation linear?

BGD

7, C4368–C4385, 2010

Authors: DEM-derived variables will be changed in the revised manuscript to SRTM DEM. We have used Pearson correlation to obtain an understanding of the spectral nature of the field data by analyzing the linear relationship between image derived variables and field data (factor axis), individually. This was commented in the revised manuscript.

R#2: Section 6.3: I would appreciate a table with the RS variables and DEM variables

Authors: We added a table with the RS variables and DEM variables

R#2: 18: Was stepwise regression used? What is a best-subset regression? Does the 'best'exist?

Authors: Stepwise regression assumes that there is a single best subset of predictor variables in the model. Unlikely, best subset regression indicates among all subset of explanatory variables, that one that does an adequate job of prediction (James and Mcculloch 1990). This is evaluated by comparing all- subset regressions, with all possible combination of predictor variables, in respect to R2, C-p statistic and number of predictors. This has been explained in more detail in the revised manuscript.

6903

Mapping plant communities

R#2: 24: change accurate into 'detailed'

Authors: This was changed in the revised manuscript.

6904

Mapping plant communities

R#2: 17: would change rapidly into clearly

Authors: This was changed in the revised manuscript.

R#2: 24: What is a most accurate map? What do you use as a reference?

BGD

7, C4368–C4385, 2010

Authors: Mean error of the predicted scores provides an indication of the overall accuracy of each kriged map. The lower this average is, the higher the map accuracy might be. This was explained in more detail in the revised manuscript.

6905

Mapping plant communities

R#2: 2: points are also spatial

Authors: This was changed in the revised manuscript to: In the final part of this procedure, score maps are associated with vegetation classes to create the final map of plant community.

R#2: 7: which scores – specify!

Authors: We specified that: . . .the predicted score values of each Factor axes.

6906

Mapping plant communities

R#2: 9: change ‘make’ into ‘making’

Authors: This was changed in the revised manuscript.

Figures

6908

Flood-duration

R#2: 6: it seems that there is also a more detailed DEM; why did the authors not use that particular DEM?

Authors: There is no more detailed DEM data available. The 40-m resolution DEM resulted from the kriging based on the 81 elevation points and the SRTM DEM provided the more detailed DEM data.

BGD

7, C4368–C4385, 2010

6909

Discussion

R#2: 14: I would start with the main aims and frame the discussion around the 4 aims; The first sentence of discussion could be moved to the conclusion section which is currently part of the discussion.

Authors: We created a conclusion section to provide the main findings of our study and provide suggestions for future studies; We started the discussion section highlighting the main aims.

6910

Discussion

R#2: 8: sentence unclear to me

Authors: This was changed in the revised manuscript to: However, even though we recognize that plan community is a spatial concept rather than a well-delimited entity (Austin and Smith 1989), we believe that the abstract definition of crisp boundaries is somehow needed to interpret space-time vegetation patterns over large areas.

R#2: 19: did all these authors use RS analysis in their studies?

Authors: We removed most of the references, leaving just Pfeffer et al. 2003

6911

Discussion

R#2: 14: change to 'for improvement in mapping'

Authors: This was changed in the revised manuscript.

R#2: 17: what is a better sample design?

Authors: This was changed in the revised manuscript to: Increasing the number of

BGD

7, C4368–C4385, 2010

samples and changing sample configuration are the most. . .

R#2: Use Wesseling et al , 1996 instead of conference paper

Authors: This was changed in the revised manuscript.

6923

R#2: Table 3: What is canopy topography? Is it the elevation that is measured with the SRTM DEM? If yes, specify that in the data section. What is *P?

Authors: Yes, it is elevation measured with SRTM DEM. *P indicates the level of significance of the correlations among variables. This was specified in the revised manuscript.

6925

R#2: Legend of DEM – is it important to choose decimals when making elevation classes?

Authors: No, in this figure it is not needed, but we would like to stress that in floodplain wetlands very small differences in elevation may cause great changes in plant communities. We left out the decimals in the figure.

6926

R#2: Figure 2: Chart of procedure: Would change 'transformation' into 'processing'; specify type of clustering; change image derived attributes into 'continuous field variables'; change accuracy of Mapping into validation. I guess that comparison is the correlation analysis; then it could also named correlation.

Authors: This was changed in the revised manuscript.

6927

R#2: Figure 3: the line 250 does not make sense, because the length of the whole trail is 250 m and not the width of starting point compared to end point.

BGD

7, C4368–C4385, 2010

Authors: The position of the line was changed in the revised manuscript.

Interactive comment on Biogeosciences Discuss., 7, 6889, 2010.

BGD

7, C4368–C4385, 2010

Interactive Comment