Hydrol. Earth Syst. Sci. Discuss., 4, S1958–S1961, 2008
www.hydrol-earth-syst-sci-discuss.net/4/S1958/2008/

Hydrology and Earth System Sciences Discussions

Interactive comment on "Modelling groundwater-dependent vegetation patterns using ensemble learning" *by* J. Peters et al.

J. Peters et al.

Received and published: 22 January 2008

Interactive comment on “Modelling groundwater-dependent vegetation patterns using ensemble learning” by J. Peters et al.

Anonymous Referee #2 Received and published: 10 January 2008

General comments The proposed manuscript is focused on evaluation of the "random forest" ensemble learning technique to model vegetation patterns at local scale. The relationship between vegetation distribution and environmental gradients is an important topic in the framework of eco-hydrology. The performed tests (classification accuracy, relevance of variables, etc.) add valuable information on the use of random forest model at local scale. On the whole, the work is well structured, the abstract is

Full Screen / Esc

Printer-friendly Version

Interactive Discussion

Discussion Paper

sufficiently descriptive and the language is appropriate and concise. The manuscript is suitable for the publication in HESS journal after some minor revisions.

Specific comments

Section 2 Since the introduction of the second study site is spread throughout the text, it is better to introduce also the "Snoekengracht" site in the "Description of the study area". It could be useful to understand the differences between the two sites. As an example, as it can be derived from Table 1 there is a different grid sampling: 20m for Doode Berme and 10m for Snoekengracht. REPLY: To my feeling, the addition of Snoekengracht at the beginning of the site description would complicate things needlessly. Therefore preference is given to first describe the Doode Bemde and discuss the monitoring strategy in detail. Then, a separate subsection, introduces Snoekengracht shortly. However, some additional information (basically stating that both sites are similar) on the Snoekengracht might be useful (as the referee indicates) and is added to the text.

Pag.3691 lines 16-23: Please verify and better explain comments on Figure 3, in particular on Figure 3b. REPLY: Lines 16-23 are rewritten.

Section 4-5 Since the misclassified pixels having the highest probability classes are mainly located around the piezometer A that shows the highest temporal variability, it would be useful to consider at least the seasonal groundwater depth in order to improve the model performances. Did the authors already perform some tests on the seasonal mean? REPLY: This valuable comment corresponds to the editor’s first comment. Valuable information on groundwater dynamics (variance, amplitude) is ignored in this study, only the average groundwater depth is taken into account. The reason for this is as banal as obvious: we did not have access to the entire time series data at the test location Snoekengracht. Therefore we could not include any information on groundwater depth fluctuations as explanatory variable in this study, since it would not allow for model testing on a spatially independent data set.

HESSD

4, S1958–S1961, 2008

Interactive Comment

Please explain the difference of the obtained results for the entire dataset Lev (Snoekengracht) on pag 3697 line 16 (overall accuracy 19.8%) and on pag. 3699 line 6 (overall accuracy 76.6%). Is the second one not trained on L? REPLY: Indeed, the objectives differ. On page 3697 line 16, results of the applicability of a random forest model trained on Doode Bemde data and tested on an independent, spatially distant data set are presented. On page 3698-3699 an assessment of a simplified ecohydrological monitoring scheme is made. A reduced data set only including important and easily measurable variables is used for model training and testing at the Snoekengracht. To make this clearer, a textual addition is made to line 2 on page 3699.

For the selection of the reduced input variables the authors did not follow the importance variable ranking. Please add more comments on the proposed choice. Moreover, to better understand the validity of the selected variables, it would be useful to add accuracy results also for the first data set (L). REPLY: Indeed, the selection is only based partly on the calculated variable importance, partly on practical considerations. The variable importance ranking is taken as a guideline to (i) select some of the more important variables, and (ii) to include an appropriate number of independent variables (oob error stabilizes). A choice was to be made to assess the model performances on Lev reduced to the 4 most important variables, or to take another important monitoring criterion (monitoring cost) into account (and increase the number of variables, since the ones chosen are not the most important ones). Preference is given to the latter, despite subjectivity. The text is changed to motivate this choice. However, if the referee finds this too subjective, the section can be changed to include only the most important variables. As the referee suggests, results for the first data set (L) are added. It is important to note that these evaluation statistics are based on the out-of-bag classification results. Cross-validation is inappropriate here because of possible bias between the variable importances of the different cross validation parts.

Finally, in order to make easier the comparison of the obtained results for the two sites and for the entire and reduced (5 variables) datasets, it is helpful to add a summarizing

HESSD

4, S1958–S1961, 2008

Interactive Comment

table including the Overall accuracy (or better the k-Cohen coefficient to also evaluate omission and commission errors) and AUC values. REPLY: We believe the addition of panel (c) to Fig. 8, and the addition of kappa values in lines 3 and 6 on page 3699 makes comparison easier. A comparison between models based on L and Lev, however, is not really appropriate in this section since model statistics are based on oob test in the case of L, and based on cross-validation in the case of Lev.

Pag.3695 line 5: Please check "true possitive rate" REPLY: Indeed, possitive should be positive.

Interactive comment on Hydrol. Earth Syst. Sci. Discuss., 4, 3687, 2007.

HESSD

4, S1958–S1961, 2008

Interactive Comment