Hydrology and Earth System Sciences Discussions

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 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.

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

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?

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?

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).

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 table including the Overall accuracy (or better the k-Cohen coefficient to also evaluate omission and commission errors) and AUC values.

Pag.3695 line 5: Please check "true possitive rate"

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

Interactive
Comment

Full Screen / Esc

Printer-friendly Version

Interactive Discussion

Discussion Paper