

Interactive comment on “Assessment of hydrological and seasonal controls over the nitrate flushing from a forested watershed using a data mining technique” by S. Rusjan and M. Mikoš

Anonymous Referee #3

Received and published: 2 January 2008

General comments

The manuscript by Rusjan and Mikos "Assessment of hydrological and seasonal controls over the nitrate flushing from a forested watershed using a data mining technique" describes the dynamics of nitrate concentration in stream water from a small watershed. A wealth of data are interpreted using a regression tree, which is claimed to help understand the factors driving nitrate flushing. On the good side, the manuscript is concise and generally clearly written, and provides a honest assessment of the usefulness and limitation of the regression tree technique in this watershed-scale biogeochemical

Full Screen / Esc

Printer-friendly Version

Interactive Discussion

Discussion Paper

application. However, the manuscript is not acceptable for publication as is, since it would need important clarifications and more detailed discussions regarding both the regression approach and the results. For these reasons I suggest a major revision.

Specific comments

- The manuscript presents a regression model to estimate the nitrate concentration. However, other more mechanistic approaches are already available (process-based hydrologic and biogeochemical models), but have not been fully acknowledged in the Introduction and Discussion sections. Such mechanistic models have been validated in applications similar to the one presented in this work and a direct comparison of the two approaches may be considered as a future development (see last paragraph in the Conclusion section).

- Some ambiguities regarding the advantages of regression trees should be addressed. These models are claimed not to require any intervention by the user (p 4217, line 18); however, the importance of the choice of the splitting values is later acknowledged (p 4218, lines 4-5). It would be thus important to extend the comments in lines 7-8 and describe whether the splitting method requires any user input or it is only dependent on the data. The discussion on page 4221 seems to suggest that the splitting values have indeed a physical meaning, but again it is not clear if they are selected by the user or not. The physical interpretation of the temperature split values should also be more carefully addressed. Are the selected temperature thresholds correlated with actual seasonal transitions (e.g., plant phenology), and do they similarly characterize the spring-to-summer and summer-to-fall shifts?

- In order to ease the interpretation of the temporal evolution of the variables, figure 2 could be changed. If seasonality is important for the dynamics of nitrate in stream water, figure 2 can be improved by adding an upper panel with rainfall and temperature plotted against "real" time, instead of "instance no.", and throughout the year. Properly placed brackets could then guide the reader to bottom panels showing the selected

Full Screen / Esc

Printer-friendly Version

Interactive Discussion

Discussion Paper

hydrologic events where also nitrate observations are available (i.e., the current figure 2). Somewhere in the text, some comments on the representativeness of the year 2006 with respect to the historic hydro-climatic averages (p 4216) could be added. Also, a clear separation between the individual hydrographs in all the figures involving the instance nos. would be useful (e.g., a vertical line).

- One of the main limitations of the proposed method is to underestimate elevated nitrate concentrations. Neglecting for a moment overfitting issues, can this limitation be overcome by decreasing the pruning factor? This option is discussed on p 4224, but I could not find any statement about the actual resulting improvements (if any). My concern here is that maybe the model is not including some attributes acting mainly during the November events. In fact, the same branches of the regression are used to characterize both spring and fall periods (figure 5), while there are strong differences between these two seasons from a biogeochemical perspective (e.g., litter dynamics). Alternatively, the proposed linear regression model may be insufficient to account for the nonlinear interaction between temperature and moisture which controls mineralization and nitrification. More comments on these or other possible explanations could be included. It would also be interesting to further comment on the prediction improvements possible with an extended dataset, covering also the rainy winter months (p 4224, lines 18-19).

Technical corrections

- Besides the following issues, a careful grammar and syntax revision is necessary.
- p 4212, line 6: change "presented" with, e.g., "is", or rewrite the sentence.
- throughout the text: please consider changing "robust" with an other term. "Robust" is generally used to positively characterize the behavior of a model.
- throughout the text: change the term "oscillation" when used with the meaning of "pulse", "peak", or "flush". "Oscillation" suggests a regular behavior that is not apparent

[Full Screen / Esc](#)[Printer-friendly Version](#)[Interactive Discussion](#)[Discussion Paper](#)

in the data presented.

- p 4215, last sentence: an estimate of soil hydraulic conductivity would be useful here, to be compared with the bedrock conductivity.

- p 4216, lines 10-11: latin names should be italicized.

- p 4217: the description of "grab water samples" is useless unless such data are showed or discussed in the following.

- p 4217, line 12: are the piecewise linear functions also continuous? If not, i.e., there are discontinuities in the dependent variable between two different subsets, are such "jumps" important, or do they mirror an actual, physical discontinuity?

- p 4218: no reference to figures 3 and 4 is made; thus, it is not clear whether the accuracies shown in these figures are the averages of the 10 cross-validation experiments. Are the regression equations reported in Tab. 2 derived from the whole data set or from one of such cross-validation experiments in particular? Please, comment on this.

- p 4219, line 8: "selections" instead of "selection".

- p 4219, lines 15-22: these sentences are not clear, please rewrite them.

- throughout the text and table 2: check consistency in variable names; in particular, check the use of subscripts and italics. Also, symbols like correlation coefficient "r" (p 4220, line 15) should be italicized.

- Table 2: a graphical representation of the regression tree (with actual branches) would be more useful than the left column in table 2.

- p 4221, line 29: indices accounting for long previous periods (e.g., avgT14, API14) can be computed only for a limited number of instances, given the brevity of the recorded hydrographs. Does this affect the role played by such variables in the regression tree? How is this accounted for in the regression?

Full Screen / Esc

Printer-friendly Version

Interactive Discussion

Discussion Paper

- p 4222, line 28: not clear. Possibly, "and simulate" instead of "which tend to simulate".
- p 4223, line 14: "address" needs to be changed.
- p 4223, line 21: "against" instead of "as regards to".
- p 4225, line 21: the comma after "states" may be removed.

Interactive comment on Hydrol. Earth Syst. Sci. Discuss., 4, 4211, 2007.

Full Screen / Esc

Printer-friendly Version

Interactive Discussion

Discussion Paper