Hydrol. Earth Syst. Sci. Discuss., 4, S2164–S2168, 2008
www.hydrol-earth-syst-sci-discuss.net/4/S2164/2008/

Hydrology and Earth System Sciences Discussions

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

S. Rusjan and M. Mikoš

Received and published: 8 February 2008

- Other more mechanistic approaches (process-based and conceptual models) available for the modeling of the nitrate exports from the watersheds are included, presented and discussed in the Introduction and Discussion section; 5 references were added. In the Conclusion section the possibility of combining the two approaches was addressed. *Further comments*: Temporally, the calculations in the mentioned models are usually performed at daily or even longer time steps; therefore, their application for the modelling of hydrologically, event-induced nitrate mobilization is limited.

- p 4217; p 4218: The comments on the regression tree construction were extended.

Further comments: The used data mining tool, the algorithm M5 which constructs the regression trees models enables automatic sieving through the given dataset. The algorithm automatically finds the most discriminating attributes which are positioned into the nodes.

- p 4221: The explanation about the selection of the splitting values was extended and clarified. The splitting values which have a physical meaning were selected by the algorithm automatically.
Further comments: The selected temperature thresholds are important from the perspective of seasonal shifts in the mineralization and nitrification found in the literature (references in the orig. manuscript) and could also present the seasonal transition in vegetation activity.
Text added (page 4221; section 4.2): For further interpretation additional vegetation characteristics and detailed description of seasonal changes in soil characteristics would have to be considered. However, as in our case, such information is usually not available, also in the mechanistic approaches these characteristics are addressed differently. This is also beyond the mainly hydrological framework of the study.

- Fig. 2 was improved. Fig. 2(A)(real time scale + vertical lines) represents the recorded rainfall events on the continuous timescale; Fig. 2(B)(instance Nos. + vertical lines) gives a detailed presentation of streamwater nitrate concentrations during the observed events.

- p 4216: Some additional information about the hydrological and climatic characteristics of year 2006 was provided.

- p 4224: Regarding the underestimation of the elevated nitrate concentration, further explanations were added.
Further comments: The main intervention of the user in the application of the algorithm is the definition of the pruning factor, the threshold for the minimum number of instances which are classified into a particular regression tree leaf. Based on other

HESSD

4, S2164–S2168, 2008

applications of the regression tree models found in the literature and also evident from our application, low threshold, i.e. small minimum number of instances in the leaves, means allowing the algorithm to overfit to a given dataset; in the sense of nonlinear interactions between the attributes we allow the algorithm to better describe the non-linearity by increasing the number of the linear regression functions. In our case, if we decrease the pruning factor we slightly improve the prediction of the resulting regression tree model (Fig. 4), however, the regression tree becomes much larger and complicated (Fig. 3). Generally, the prediction of the regression tree with 125 instances in the leaves is respectably good even during the highly variable nitrate concentrations in spring and summer rainfall events. The main problem is the single recorded hydrograph in November which caused an extreme increase in the concentration, but on the other hand, this hydrograph is also a good example of some limitations in the regression tree algorithm application. To be able to acquire more accurate predictions during the autumn rainfall events, the only real solution is extending the dataset with more autumn hydrological event observations, as it is the case for spring and summer rainfall events.

- p. 4221: Regarding branches in Fig. 6 (before Fig. 5): It is true that the same branch is used to characterize the spring and autumn periods on the second split level of the dataset (Fig. 6). However, if we look further down the regression tree (new Fig. 5), the temperature avgT1 on the third split level separates the spring period from the autumn period (linear models Nos. 8 and 9 in Table 2).

TECHNICAL CORRECTIONS:

- p 4212, line 6: Corrected.
- The terms *robust* and *oscillation* were changed where appropriate.
- p 4215: The data about the soil hydraulic conductivity were added.
- Latin names and attribute marks were italicized throughout the text.
- p 4217: Grab samples were mentioned to stress the applied approach where the multi-parameter sonde was not only regularly calibrated but also compared to real field

HESSD
4, S2164–S2168, 2008

values defined by laboratory analysis. Statistical comparison of the measurements was added.
- p 4217, line 12: The linear functions in the regression trees are piece-wise and by definition discontinuous but the algorithm considers the statistical thresholds that mediate and smooth the jumps between the functions. If the discontinuities in the data are too large they cannot be mediated and vice versa, it is not necessary that the actual jump between the function means an actual physical discontinuity (details can be found in the references in the manuscript).
- p 4218; Fig. 3 and Fig. 4: The regression tree derivation performing a 10-fold cross-validation is explained under Methods/Regression trees section and the caption of both figures was extended by mentioning this method.
- p 4221: The comment on the derivation of the linear equations reported in Table 2 was extended. The cross-validation method application considers all parts of the split dataset.
- p 4219, line 8: Corrected.
- p 4219, lines 15-22: The sentences were rewritten.
- The consistency of the variable name was checked.
- Table 2 was changed; the schematic representation of the regression tree model was added and is shown in new Fig. 5.
- The visibility of figures Nos. 2, 6 (original Fig. 5), 8 (original Fig. 7) was improved by adding vertical lines which separate the periods of the observed rainfall events.
- p. 4221, line 29: The indices which are used as the attributes for the description of the preceding seasonal biogeochemical and hydrological conditions were calculated for the whole period of the year 2006 based on continuous monitoring of hydro-meteorological conditions and not only for periods of the recorded hydrographs.
- p 4222, line 28: Corrected.
- p 4223, line 14: Corrected.
- p 4223, line 21: Corrected.
- p 4225, line 21: Corrected.

HESSD
4, S2164–S2168, 2008

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

HESSD

4, S2164–S2168, 2008

Interactive Comment