The article is interesting to read, but seems to be lacking some information that is necessary for better understanding by the reader. I suggest that the authors describe certain parts more clearly and consider setting more subtitles to keep different parts of model description and results apart.

A comparison or discussion on pros and cons of the used model for retention in comparison to other possible models would be interesting.

Some questions to be clarified:

a) It is stated clearly which inputs are used, but the model description is confusing. Which parameters are estimated? Are all parameters areas specific, and if so do they vary a lot between areas? How is expert knowledge used in the fitting of the model. E.g. for equation 1 are there parameters estimated in all parts of this formuls (S, P, D and R) or are some of them observed or considered known. This information is given in the text later, should however be given right after formula 1 (e.g. page 10836 line 14 states what is assumed to be known, move this ahead).

b) The total loss (S) is modelled from 3 land cover classes (cultivated, wetlands and other land). Do these 3 land cover classes add up to 100% of land cover? If so this should influence the estimation of the 3 parameters, since the variables will be linearly correlated. How is this handled? If there are land cover classes not in the model, this should be stated clearly.

c) Two formulas are given to compute/estimate retention. I the difference between them that one is used if there are lakes in the area, whereas the other one is used if there are no lakes? Or how do you choose between these for the different basins? Is lambda the same in these two models, i.e. if lambda a common estimate for both equations? State in the article. Hesse et al. ECOLOGICAL MODELLING Volume: 269 Pages: 70-85 made comparisons for different retention models. This might be interesting for you to comment in the article.

d) The risk of overfitting/overparametrisation is mentioned and given as reason that retention parameters are the same for all source categories. Is this reasonable and can be motivated? How? How do you control for overfitting in this model, is it by only allowing a few parameters to vary or do you control it? Would any kind of cross-validation help to avoid overfitting?

e) In page 10837 line 9 you talk about the total N retention that is estimated. Does this regard fitting $R*S_i + R*P_i + R*D_i$, related to equation 1? When you do fitting on different groups, are parameter estimated individually for a group? If 10 danish subbasins form one group, how many parameters do you estimated from those, is it 4 (3 theta and 1 lambda) or more? Are estimates for thetas and lambda very different for the groups of basins? Parameter estimates should be given, at least as example.

f) If groupings of basins is made due to geographical location or similarities, would not that suggest dependence/correlation between the basins and influence p-values (with the concept of statistical inference based on independent observations). The error term in (1) does not indicate that dependencies are taken into account. Can p-values be trusted?

g) In the results unit-area specific loads are discussed. As the model is designed to predict N load rather than unit-area loads: was this expected? Could the model be adjusted if unit-area loads are interesting? Could this be a result of overfitting in the original model?

h) In figure 4 the relationship between estimated retention and total drainage area are given. In these figures it seems that drainage area has no influence on retention in %, whereas lake area (%) has a clear nonlinear relationship. How do these curves related to equations 3 and 4? Probably the equations and estimated parameter lambda are used to compute the estimated retention, i.e. the curves should reflect the relation in 3 and 4. Is this true? The line shown in the plot 'retention and lake area', why is it plotted there? How is it related to the model? Since this line does not fit well, does this indicate that the model does not fit well?

i) Also the function fitted to specific load and lake area (%) is strange, why do you use this fitted line instead of an exponential/logarithmic relationsship or a square-root relationship. Where does the function come from? How is it motivated?

Smaller notes

Relative differences are used to give equal weights to small and large basins. A motivation why this is a good choice in this context would be appreciated.