Biogeosciences Discuss., 7, C854–C856, 2010
www.biogeosciences-discuss.net/7/C854/2010/

Biogeosciences Discussions

Interactive comment on "A linear mixed model, with non-stationary mean and covariance, for soil potassium based on gamma radiometry" *by* K. A. Haskard et al.

Anonymous Referee #1

Received and published: 5 May 2010

In this paper, a set of linear mixed models developed by same authors previously are applied to model soil potassium content as a function of the passive gamma-ray emissions of the earth surface. The performance of the traditional stationary variance model is compared to new proposed non-stationary models, the emphasis being on a detailed description of the different statistical models and the tools used for model selection. This is a well written paper targeting a specialized statistical audience, being a valuable contribution in the field of statistical methods applied to soil science. Below are some comments that the authors might want to take into account.

Comment 1: This is a minor comment however important: the line numbering is missing

which makes the reviewing process difficult.

Comment 2: The authors made the point that the two model selection tools AIC and prediction error do not agree in terms of preferred model. This is a common problem, occurring very often in statistics, and depending on whether or not the model is used for prediction or as a descriptive model, one tool is preferred to the other. Another way to overcome this is to use the cross validation method to calculate the prediction error. What is commonly used for example is a k-fold cross validation method, which allows to select the best model based on the cross-validation error.

Comment3: Is there a particular reason why the training and validation samples were so unbalanced (222 against 661 observations), the common practice is to have half and half, when the third "test" sample is not used.

Comment 4: In terms of model assessment, in addition to the AIC and the different prediction errors presented in table 1, the non-statistical audience would appreciate better the value of R2, or R2 adjusted as a measure of model fit. In addition a plot of the observed against fitted values would be very useful to visualise the performance of the different models (or at least of the preferred model).

Comment 5: The statistical model described in section 2 would benefit of some references, since not all the formulae presented are straightforward (eg. Formula of matrix P at the bottom of the page3, first column and formulae 5 and 6).

The remaining are minor comments Comment 6: In the abstract the following sentence: Non-stationary models are proposed. . .., is missing somehow the verb, needs to be rewritten. Same comment for another sentence in the introduction (page 2, the bottom of first column): A covariate which has been . . ...

Comment 7: Small z should be used instead of capital Z in the formula of u_p at the bottom of page 3, column 1. Same at the bottom of page 5, small z instead of capital Z should be used for consistency.

Comment 8: Past tense should be used when reporting the results in the results section.

Comment 9: “the” appears twice in the second sentence of the results section.

Comment 10: The sentence at the bottom of page 6 “ the use of a separate validation data could be used. . ...” needs to be rewritten.

Interactive comment on Biogeosciences Discuss., 7, 1839, 2010.