## Authors reply to the comments of the reviewer

**Title:** The application of GIS based decision-tree models for generating the spatial distribution of hydromorphic organic landscapes in relation to digital terrain data

**Authors:** R. Bou Kheir, P. K. Bøcher, M. B. Greve, and M. H. Greve

**Hydrology and Earth System Sciences Discussion, 7, 389, 2010**

**Dear Dr. Hengl,**

**You will find enclosed the corrected version of the manuscript. We would like to thank you for the time and efforts spent on the paper; which we believe, have improved the manuscript by making it more concise and clearer. To meet the reviewer comments, corrections were carried out as follows.**

#1 On P395L17-19 I can see that the authors have decided to convert the original SOC values to 2 categories (<10%; >10%). By doing this, the authors threw away a lot of important information (e.g. about variable distribution; location of possible hot-spots etc) from the analysis. Why? I hope that there is a good explanation for this, otherwise I would suggest that the authors run the analysis with the original variable (SOC in % or even better kg/mˆ3 of soil) and then specify the cross-validation results in the original scale (which could be less optimistic than the current 75% accuracy).

**We want to classify the wetland soils in the studied area as "organic" and "not organic", a categorical decision for land use management (see page 393, lines 20 to 25); this is the reason for using classification trees. In addition most of the existing field surveys (on which depends our study) are providing categorical values (mineral or organic soils): 85% of the existing field samples are categorical data (mineral or organic soils). For the samples having organic matter as continuous value, the range is between 0.2% and 96.6%.**

**A sentence was added in the introduction as follows:**

**The purpose of this study is to implement CTA and evaluate its ability to provide accurate soil landscape prediction and more precisely to determine the geographic distribution of hydromorphic organic landscapes (target variable being the soil organic carbon classified as organic or mineral soils depending on the available data collected during the last 60 years) at an unsampled area in Denmark from mapped environmental variables.**

#2 My problem with using classification trees for spatial prediction is that this method completely ignores spatial locations of point samples (see also Henderson et al., 2004, pp.394–396). It is not clear from this article how did the authors worked around this problem and what would be their remedy. If the SOC values are spatially autocorrelated (which I assume is highly possible from Fig.3), then the model estimates is biased in the areas where the points are more clustered. This makes this method statistically sub-optimal to geostatistical techniques such as regression-kriging, GWR or BME.

**We are working actually on comparing the indicator krigging and the decision trees for a predictive mapping of clayey soils in Denmark. The obtained results (through validation) indicate that the accuracy of maps obtained by applying decision trees is higher than those produced by indicator Krigging. In all cases, the application of a large number of classification trees (186) in the current work with different environmental parameters can be considered as a remedy for the clustering of sample points. We are testing different environmental parameters at each time (for each tree), whatever the spatial locations of point samples. Decision-trees can reduce the need for extensive sampling and costly laboratory analysis by minimizing the number of samples needed to generate spatial prediction.**

**Other minor corrections:**

1. P390L19: incomplete "... was the combined...";
**Done**

2. P392L23-25: what about uncertainty - can it include information on the uncertainty of estimates?
**Of course, there is some uncertainty, and the accuracy of the classification trees will differ according to the input predictor environmental parameters; for that, a large number of classification trees should be constructed to minimize such uncertainty.**
**A sentence was added in the introduction as follows:**
**'However, in the built classification trees, the uncertainties of the classes in each one of their leaves can be noticed. The good behavior of testing several trees (with different predictor environmental parameters) could minimize such uncertainties to a wide extent'.**

3. P395L5: show the location of samples in Fig. 1.
**Done**

4. P395L17-19: why not use the original variable?
**Most of the existing field surveys (on which depends our study) are providing categorical values (mineral or organic soils): 85% of the existing field samples are categorical data (mineral or organic soils).**

5. 2.1 I would appreciate in this section a histogram of the target variable and/or a bubble plot of values (spatial spreading of sampled values);
**Most of the existing field surveys (on which depends our study) are providing categorical values (mineral or organic soils), and therefore it is not possible to draw such histogram**

6. P398L18: if existing, add a reference "special", otherwise claim a new method that your group developed;
**Two references were added about the special hydrological algorithms (Tarboton et al., 1991; Chorowicz et al., 1992).**

**Tarboton, D.G., Bras, R.L., and Rodriguez-Iturbe, I.: On the extraction of channel networks from digital elevation data, Hydrological Processes, 5, 81-100, 1991.**

**Chorowicz, J., Ichoku, C., Riazanoff, S., Kim, Y-J., and Cervelle, B.: A combined algorithm for automated drainage network extraction, Water Resour. Res., 28(5), 1293-1302, 1992.**

7. 3.2.2 Free and open source GIS SAGA can derive 2-3 times more DEM parameters than ArcGIS 3D analyst; including an iterative TWI.
**The free and open source GIS SAGA is of high interest, and will be used in the near future for extracting DEM parameters.**

8. P399L20: this is not a good argument; I would assume that climatic conditions are rather homogeneous because there is not much landscape in Denmark. A variety of climatic images (MODIS, worldclim.org, Meteosat) are available for free.

**The sentence shown on page 399 – Line 20 'We did not use climatic data in this study because meteorological stations were rather few. Moreover, the main factor controlling climate in this region is elevation, which was retained in the constructed classification tree-models' was replaced by the following sentence:**
**'We did not use climatic data in this study since Denmark is a relative small country with low topographic relief. Moreover, the main factor controlling soil moisture is local topography and soil conditions, which were retained in the constructed classification tree-models'.**

9. P400L4-5: by "assign" I assume you mean "overlay".
**The word 'assign' was replaced by 'overlay'**

10. P401L22: "was obtained under a GIS environment" - please provide more detail about the processing steps; Fig. 1: show location of all sampling points and/or bubble plot of values.
**The clarifications were given as follows:**
**Using the resulting preferred classification tree-model (having the highest predictive power, and the lowest number of terminal nodes and predictor parameters), a predictive map of SOC was obtained under a GIS environment through the application of the prediction classification tree rules (shown in Fig. 2).**

References of interest:
- Henderson, B. L., Bui, E. N., Moran, C. J., Simon, D. A. P., 2004. Australia-wide predictions of soil properties using decision trees. Geoderma 124 (3-4): 383–398.
- Hengl, T. 2009. A Practical Guide to Geostatistical Mapping, 2nd Edt. University of Amsterdam, 291 p. ISBN 978-90-9024981-0 [http://spatial-analyst.net/book/GstatIntro]
**The references were added on the modified version of the manuscript (as requested).**