**Hydrology and
Earth System
Sciences
Discussions**

Interactive comment on "The application of GIS based decision-tree models for generating the spatial distribution of hydromorphic organic landscapes in relation to digital terrain data" *by* R. Bou Kheir et al.

T. Hengl (Referee)

t.hengl@uva.nl

Received and published: 15 March 2010

This is a well written and easy to follow paper on the application of Digital Soil Mapping techniques for mapping of organic landscapes in Denmark. Although the novelty of methods presented is limited (similar methods have been already described in e.g. McKenzie and Ryan (1999) and Henderson et al., (2004)), I would welcome its publication in your journal provided that some clarification is added in the revised version of the article.

#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).

#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 auto-correlated (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.

Other minor corrections:

1. P390L19: incomplete "... was the combined..."; 2. P392L23-25: what about uncertainty - can it include information on the uncertainty of estimates? 3. P395L5: show the location of samples in Fig. 1. 4. P395L17-19: why not use the original variable? 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); 6. P398L18: if existing, add a reference "special", otherwise claim a new method that your group developed; 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. 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. 9. P400L4-5: by "assign" I assume you mean "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.

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]

Interactive comment on Hydrol. Earth Syst. Sci. Discuss., 7, 389, 2010.