**Geoscientific
Model Development
Discussions**

Interactive comment on "Sensitivity analysis and calibration of a soil carbon model (SoilGen2) in two contrasting loess forest soils" *by* Y. Y. Yu et al.

Anonymous Referee #1

Received and published: 10 September 2012

Dear all Presenting the development and testing of a model is always a challenge in terms of writing a paper. Authors here find the right way to do so. The paper is well-written and gives a good overview of the issue. I have few concerns that might only require clarification, supplementary details and/or one or two more figures.

1- as mentioned in the title, this paper presents sensitivity analysis and calibration of a model but not the model validation. Nevertheless I think a model has a value only if it is able to reproduce the reality. I would have appreciated to find a "model validation" in this article. Authors work on 3 pedons per site. If they do not have access to other data (which is unlikely), several solutions are available to them to present if not a true validation, at least a stopgap: i- working on 2 pedons to calibrate the model and validating it on the third one. Repeating this process on another couple of pedons and

so on, to discuss the difference between simulated and observed data in each of the 3 cases. ii- at least simulating each pedons and not only the weighted average of the 3 pedons

2- Belgian pedons are for the 3 of them cutanic fragic Albeluvisol (3 orders of precision) whereas Chinese pedons are for 2 of them Kastanozem (1 order of precision) and for the last one a Luvisol. On which type of soil each calibrated set of parameters can be used for? What is the range of application? by only looking at soil description, it seems that the Belgian set of calibrated parameters are restricted to very specific application whereas the Chinese one could be used for quite different types of soil (Kastanozem and Luvisol have not similar properties). Independent presentation of each pedon would perhaps contribute to address this issue

3- by looking at the 2.5 "calibration approach" part, reader could expect numerous run to get the right and precisely defined values. He could be disappointed by finding only 14 and 8 runs in the "3.2 - calibration" part. Could authors explain why they can minimize each function (representing important parameter) in so few runs?

4- RMSE are very frequent in calibration methods. That's not the case for MD and DIS. Could authors explain the added-values of MD and DIS over RMSE?

5- I did not catch why the calibration in China accounted only for 2 parameters and not for 3 as in Belgium. KRPM remain at 0.3 for all tests.