

## ***Interactive comment on “Underestimation of boreal soil carbon stocks by mathematical soil carbon models linked to soil nutrient status” by B. Ľupek et al.***

**Anonymous Referee #2**

Received and published: 7 March 2016

Review of Tupek et al.

Summary:

Three soil models (Q, Yasso07 and CENTURY) are ran against Swedish forest soil inventory data to gauge how well they can estimate soil C stocks. The soils were additionally broken down into 10 distinct groupings based on soil characteristics or 5 on site characteristics. Generally the models perform well enough but have problems with certain sites characterized by high fertility and are generally well-sorted for parent material.

I have some troubles with understanding the point of the paper. The authors took three

C1

separate soil C models and ran them then compared them. That is fine but why not have examined how the special characteristics of CENTURY could have helped its performance? The authors noted that CENTURY simulates its soil C to only 20 cm and they noted that it should likely be increased by 40-50% like Yasso07, but then why not show on plots how that would look? Similarly, CENTURY is capable of N dynamics and the authors explicitly note that N deposition at some sites seems to be important, so why not do a run with the N-cycle turned on? Then at least we could see how well the model does when its full capabilities are used. This strikes me as taking a Ferrari, deflating all of its tires, filling it with poor petrol and then racing against a Honda. Sure it's performance can be evaluated but it is hardly ideal conditions to see how fast it can really go.

I also worry about the litter inputs. I would have liked to see some way of independently evaluating the litter input contributions.

I recommend the authors do some further simulations to make this paper more interesting and to offer up a better analysis of how the model processes can contribute to estimated SOC stocks (thinking here the N cycle in CENTURY). I usually don't like to ask for more simulations but in this case I think it is necessary to make the paper have wider appeal. If not a more specialized journal could be appropriate.

Specific comments:

1. The paper is generally not well written and would greatly benefit from English copy-editing. I mention this as I often had to re-read sections to understand what was written. There are a few areas where I still don't understand what was being communicated.
2. The section on fAPAR was hard to follow ('actual state'? I don't understand if this was an English problem or if this term was meant. It is a strange term to be used). In the end I was not sure how good this fAPAR method worked out. I can't see anywhere that this was explicitly tested against some sort of observations. Since the litter inputs are pretty important to drive the models with, shouldn't this be very well evaluated?

C2

3. How was the stump defined for the biomass? Usually I think of stem, coarse roots, and fine roots with the stump being what is left after a site is logged. How was it meant here?
4. Line 264 - But the CENTURY simulation was run to equilibrium, right? Also how was equilibrium defined for the models?
5. Table 2, how is the productivity class derived?
6. Table 2 - The depth of soil is assumedly cut off at 1 meter?
7. Table 3 - Parameters (leftmost column)? What is meant here? How the model was parameterized? I found this confusing.
8. Table 3 - CENTURY, is the soil depth adjustable from 0.2? Could it be increased to 1.0 to more simply make it comparable to the other models?
9. Figure 2 and text in main - Soil group 8 has only 8 samples within it. Is this reasonable to keep as a group? Given how many uncertainties develop as this regression tree is created (calculation of SOC, assignment to weather stations, measurement uncertainty, etc.) is it reasonable to let a group be only 0.24% of the total?

---

Interactive comment on Biogeosciences Discuss., doi:10.5194/bg-2015-657, 2016.