Journal: Hydrology and Earth System Sciences
Title: Projected decrease in wintertime bearing capacity on different forest and soil types in Finland under a warming climate
Authors: I. Lehtonen, A. Venäläinen, M. Kämäräinen, A. Asikainen, J. Laitila, P. Anttila and H. Peltola
MS No.: hess-2017-727
MS Type: Research Article
Iteration: First review
Referee #1

We would like to thank the referee for the constructive comments and suggestions. Our replies to the comments are given in "Italics" after each specific comment.

The paper presents the results of a study aimed to evaluate the projected decrease in the bearing capacity of Finnish soils in function of the changing climate during the 21st century. The paper appears well written and the results are interesting for the scientific community, even if related to the specific territory of Finland. The method is general and can be applied also in other nations in which the wood harvesting is economically important.
However, a lack in this paper is the detail related to the choices of parameters performed in the model used, the description of the pre-processing procedures (inclusive of the choices of the several parameters used in this study), the statistical comment about the values (especially those selected as a result of several simulations), and in general a too short description about the consequences and the limitations of these choices on the interpretation of the results. In my opinion, this part deserves a deepening, because it could help to evaluate the results and also give more strength and robustness to the conclusions. This is the reason for which I do not think that this paper could be accepted in the present form, but requires some modifications that, in my view, can be intended as minor. The list of requirements can be understood better by looking at the specific comments here listed page by page.

– Introduction: in my opinion, a too large part of the introduction is dedicated to explain the industrial problems, while a too small part is dedicated to the scientific problem and the models used.

We agree that the introduction can be reorganized and additionally shortened in general.

– Page 4 lines 4-28: the equation proposed to estimate soil temperature seems not consider the effects of soil moisture (unless thermal conductivity is kept variable, but since there are no measures of soil moisture it is hard to consider such variations). A comment on this consideration may be required.

Yes, the equation assumes constant water content over time. We agree that it would be a good idea to add a comment about this in the manuscript.

– Page 5 line 15: regarding KT values, is the interval of values used significant for the considered soils?

Yes, we think so. According to the study by Rankinen et al. (2004) where the used soil temperature model was first introduced, the calibrated K_T values varied between 0.5 and 0.8 at five stations across Finland. They had calibrated K_T at 20 cm and 50 cm depths and the highest value they got (~0.8) was *at Sodankylä station, which seemed to have the highest K_T values also among our stations. The soil type at Sodankylä is sandy gravel and soil types with more fine-grained texture tend to have smaller K_T values. Jungqvist et al. (2014) used optimization interval 0…1 $Wm^{-1}K^{-1}$ for K_T but we noted that 1 $Wm^{-1}K^{-1}$ is not enough for the higher limit at deeper soil depths so we extended the higher limit to 2 $Wm^{-1}K^{-1}$ although we mainly focused on 20 cm soil depth and at that depth, the optimized values did not exceed 0.7.*

– Page 5 lines 18-19: "while, for example, KT seemed to steadily increase with soil depth." this is consistent with the assumption of increasing soil moisture at increasing depth (or change of soil texture): do you have any data evidencing these facts? Please comment.

This is undoubtedly one reason for this. For example, in the report by Soveri and Varjo (1977) cited in the manuscript, they show in Table 4 measured soil moistures from one test site over one winter season. Based on those observations, the soil moisture, on average, increases with increasing soil depth. In Table 5 they show typical heat capacity and heat conductivity values for different soil types with different soil moistures. Based on those values, the heat conductivity increases rapidly with soil moisture. In the report by Heikinheimo and Fougstedt (1992) are shown the soil textures at some of the stations on certain depths. For example, at Anjala the share of clay increases with increasing depth whereas, e.g., at Maaninka there is almost equal amount of silt and sand near the surface and also at 0.7 m depth but mainly sand around 0.5 m depth and below 0.7 m.

– Pages 4-5: the method elaborated to retrieve soil thermal conductivity is strongly linked to the availability of soil temperature data, and thus will become representative of the experimental sites during the measurement periods. If I have correctly understood, such values optimized for each site will be adopted for the following simulations. However, there is no any reason for which such values could remain constant also in future climate... This could be a limitation for the reliability of future simulations. If authors do not agree with my conclusion, they could explain why...

Yes, the parameter values were kept constant throughout the simulations after the parameters were defined for the three soil types based on the calibration period. Of course, it is clear that there are a lot of uncertainty in the parameter values and in reality they are never absolutely equal in two different places. Moreover, an almost equally good correlation between the observed and simulated soil temperatures can be achieved with very different set of parameters: if you modify one parameter, you can further modify the rest of parameters conveniently to achieve virtually still as high correlation as previously. Thus, we first set values for the least sensitive parameters and only K_T was optimized during the last phase when all the other parameters were kept constant as the results seemed to be most sensitive to K_T. At this stage, we moreover used only one station representing one soil type, so these three stations are basically representative examples for the soil types. However, as can be seen from Table S1, high R^2 values were achieved also at many other stations for different soil types, e.g., at Maaninka both for clay/silt and sandy soil. The model performed reasonably well even with the wrong soil type at many locations during the calibration period (see Table S1). As the used stations are moreover located in areas representing quite a different climatic conditions, we assume that possible changes in soil characteristics, including thermal conductivity, do not crucially change the results.

– Page 6 lines 1-10: the choice of different thresholds for soil freezing changed substantially the evaluation of the number of days with frozen soil? How and how much?

For example, at Ylistaro station the soil temperature was during the calibration period below 0 degC at 20 cm depth on 81 days and at 50 cm depth on 58 days annually, on average, when taking into account that soil temperature was observed once every fifth day. The modelled soil temperature was below 0 degC for clay/silt soil type configuration at 20 cm depth on 132 days and at 50 cm depth on 124 days annually. The modelled soil temperature was below -0.5 degC at 20 cm depth on 103 days and at 50 cm depth on 51 days annually. In addition, the observed soil temperature at 20 cm depth was also on 30 days annually exactly at 0 degC as the accuracy of soil temperature observations was 0.1 degC.

There was also huge variability in the number of days with soil temperatures below 0 degC between individual stations, evidently induced by differences in local characteristics. For example, at Apukka station soil temperature at 20 cm was only on 35 days annually below 0 degC while at Sodankylä there were 178 such days and at Lompolojänkkä the observed soil temperature had never (only three years of observations from there) been below 0 degC at any measurement depth. All of these three stations are located in Lapland between 66th and 68th northern latitude in similar climatic conditions. These three stations had very different soil characteristics but also within stations with more similar soil types there existed surprisingly notable variability in the number of frozen days.

– Page 6 lines 17-23, and page 7 line 5: I suggest to say here that the values used in eqs. 5 and 6 will be discussed later.

Ok

– Page 8 lines 3-12: again, the method elaborated to retrieve the values of parameters is strongly linked to the availability of measured data, and thus will become representative of the experimental sites during the measurement periods. Since such values optimized for each site will be adopted for the following simulations, there is no any reason for which such values could remain constant also in future climate... Also in this case, if authors have a different idea, they could explain why...

The parameters optimized at each station located in different climatic conditions across Finland were averaged over all the stations to achieve the final parameters, which were then used in validation of the snow model (except kmax and kmin related to the solar azimuth angle having the latitudinal dependence). The validation period moreover had different kind of winters, cold and mild, snowy etc. If we had used the parameter values optimized for each station, the R^2*-values would have been approximately 0.01 higher (page 8, line 15). Based on Table S2, the snow model performed equally well in different climatic conditions in different parts of Finland. However, we clearly see that the model with the optimized parameters performed worse before 1981 than thereafter, which we think is largely attributed to the correction factor for solid precipitation (cps), which we think had been higher previously due to a larger measurement error. Of course, it is not impossible that there have been some shifting in other parameters as well. On the other hand, part of the parameters are linked to things like freezing point of water or solar azimuth angle, which we can easily assume to stay constant.*

– Pages 8-9 lines 29-4: in this paper, many decisions about parameters are just summarized by “hiding” the results. For instance, in this case, the choice of values for kmin and kmax is not justified, and the reader cannot understand how it has been made. In my opinion, this may deserve an additional subsection (similarly as all other choices of this model).

We will elaborate the choice of the parameters in more detail. The impact of forest canopy for kmin and kmax was estimated based on Vehviläinen (1992).

– Page 9 lines 9-10: authors use only R2 as indicator of good simulations. However, if – just for example - I would have a simulation in which simulated snow depth has almost the same time trend of observations, but a value that is double, R2 will be close to one even if the relative error will be 200%... I suggest to use also bias or standard error as a criterion to validate simulations, and not only use correlation coefficient (and, by the way, it is better to use R and not R2).

Due to this issue, in calibrating the snow model we minimized root mean square error (p. 8, l. 8) instead of maximizing R^2. In Table S2 we show in addition to R^2 the relative error. During the calibration period, the modelled and observed snow depths are, on average, close to each other in addition to high R^2.

– Page 9 lines 30-32: how the GCM and RCM have been chosen (I think you should mention here more clearly that the detailed list of model chosen is reported in Table S3), and why those models, among the whole EURO-CORDEX dataset?

We will add details about the model choice. Basically, these GCMs were chose because we had done the bias-correction for those models in a previous project. Those models were originally chose based on their skill in simulating present-day temperature and precipitation climatology over northern Europe. The RCMs were chose because we wanted to use a set of models with a uniform bias-adjustment approach and this set of models with a uniform bias-adjustment approach had a largest number of simulations available.

– Page 10 line 20: is the modeled annual average number of days evaluated as the average of all GCMs and RCMs, respectively?

Yes, the multi-model means are shown in the figure.

– Page 12 lines 6-15: how large is the difference among model ensembles (separately for RCM and GCM) in the three climatic periods? I think that also this information is important to statistically locate your results. Section 3,4 and Figure 6, in my opinion, are not informative, as they mention only the two models giving the maximum and the minimum values, and not the distribution. As climate cannot be described just by extremes, but needs a complete statistical information, for the same reason I think that the standard deviation or some equivalent statistical parameter can be more informative about the dispersion of individual model calculations.

We partly agree and partly disagree with this comment. We think that Fig. 6 showing the complete spread among the model ensembles is informative but of course some more information about the distribution could be also added. For example, a scatter plot showing the area-averaged number of days with good bearing capacity for each individual model would illustrate the distribution very well.

– Table 3: the numerical values given for each parameter have too many digits, most of them without any statistical meaning. Instead of giving a number with too many not significant digits, authors should give a number and an error associated with the experiments and comparisons, like a = value ± error

It is hard to estimate the error associated with most of the parameters as in many cases if you change the value of one parameter completely, the model can be still adjusted to perform equally well by adjusting the values of other parameters as well. So, the parameters are intrinsically not exact at all but we have given here those values that were used in our calculations. Moreover, for example, one of the parameters is Napier's constant and we are unsure how to measure the error in its value.

– Figure 2: since it is hard to appreciate differences among the three figures, given the quite large interval of variation of the number of days, it could be better to plot, for second and third column, the differences among GCM and observations, and RCM and observations, respectively (similarly to what you did for Fig, 3). Or maybe you can add such figures, if you want to keep the total number of days.

This can be changed.