Journal: NHESS
Title: Risk for large-scale fires in boreal forests of Finland under changing climate
Authors: I. Lehtonen, A. Venäläinen, M. Kämäräinen, H. Peltola, and H. Gregow
MS No.: nhessd-3-4753-2015
MS Type: Research Article
Iteration: First review
Referee #1

We would like to thank the Referee for the in-depth review and detailed comments. Our replies to the comments are given in "Italics" after the comments given in the beginning of this document.

General comments to the author

There are several reasons why this article is not ready to be accepted.

1. The article is not very innovative. There are several studies about the impacts of climate change on the fire regime, using (different types of) models to simulate burnt area of the number of fires, based on fire weather data or indices, computed with future projections of meteorological data, simulated by one or several GCM/RCM. Some of these studies were performed for Finland in spite of, fortunately, this country being only marginally affected by wildfires. In the introduction, the authors seem to justify this paper with the assessment of the socioeconomic and ecological impacts (missing in the previous studies performed for Finland) but this pioneering aspect (not included in the objectives) was not considered in this study. The authors need to justify this paper, addressing relevant scientific and/or technical questions within the scope of NHESS and presenting new data and/or novel concepts, ideas, tools, methods or results;

We think that one of the key aspects of this manuscript is the use of five different GCMs independently which revealed the large uncertainty related to the projected change in forest fire danger only due to the uncertainty related to the magnitude of climate change signal. Those previous studies which used only one climate model (e.g., Mäkelä et al., 2014; Yang et al., 2015 for Sweden) cannot address this uncertainty very well and those studies where several models have been used, have applied usually the multi-model mean approach (e.g., Kilpeläinen et al., 2010; Lehtonen et al., 2014b; Sherstyukov and Sherstyukov, 2014 for Russia). One important shortage noted, e.g., by Lehtonen et al. (2014b) was also that applying delta change method with multi-model mean change assumes the variability of future climate to remain unchanged compared to the reference period. For example, Yang et al. (2015) recently provided spatially rather detailed projections for future forest fire danger in Sweden based only on one climate model. We believe that our study compliments these earlier studies substantially since we have shown here that current climate models produce wide range of projections for forest fire danger with a selected emission scenario, ranging from almost negligible change to a large increase. As climate variations explain only a part of the forest fire activity, it is clear that true uncertainty related to the future forest fire activity is even larger. We certainly think that this aspect in general could be better emphasized in the manuscript, perhaps partly at the expense of focusing on large forest fires.

Moreover, our results revealed interestingly that large forest fires in Finland are predominantly human-induced during late spring and early summer when forest fire activity in Finland has been recently largest. Contrary to that, lightning plays important role during midsummer when also the link between fire activity and weather is strongest. Furthermore, this means that the projected changes in fire activity are largely attributable to the changes in weather conditions during this time of year, reflected also by the weak correlation between the fire activity and weather in May when a large majority of fires are caused by human activities. Although Tanskanen and Venäläinen (2008) already had noted the early season peak in the fire activity (which contradicts the conventional understanding that most of forest fires occur in mid- and late summer) based on data over a shorter time period, they did not study the causes of fires. They only speculated that this might be due to, e.g., prescribed burnings and burning of trashes and our results confirm their hypotheses.

2. This paper is unnecessarily confusing in fundamental aspects, namely the data (both fire and climate) and methodology sections. First, the authors need weather data (air temperature and air relative humidity, wind and accumulated precipitation) at the same location and time (12:00) to compute all the components of the Canadian Fire weather system as well as their spatial/temporal means (FWI, DSR, MSR, SSR, etc.). For the sake of dataset's homogeneity, the authors should have selected just one meteorological dataset which can provide all the necessary data. Such data sets exist. For example, the ERA Interim dataset comprises all the required data. However, the authors selected two different databases: the ERA Interim for wind speed and the Finnish Meteorological Institute (FMI) observed dataset for air temperature and relative humidity, wind and accumulated precipitation. This decision has the additional drawback of the FMI dataset being composed by daily mean values of those variables which required the estimation of afternoon values, a procedure that unnecessary increases the data errors. It seems a little strange that FMI dataset comprises mean values but not the value observed at noon as well as values of air temperature, relative humidity and precipitation but not wind speed. Second, the study period is 1996 – 2014 (please replace minus sign by endash whenever defining time periods) however, the dataset was enriched with 9 extra fires, 7 of them before 1971 (almost a climatological period of 30-years before the study period). What is the need to include this small number of large fires? Why did the Finnish fire dataset include a Swedish fire? These fires increased the fire models performance or the confidence/statistical significance of the findings? How? Third, the authors decided to use high spatial resolution climate data. So, why use GCM instead of RCM simulations? This is not very important but, please explain why to choose $0.1^{\circ} \times 0.2^{\circ}$ grid and not $0.1^{\circ} \times 0.1^{\circ}$. The use of GCM outputs required the downscaling of this data. The authors decided by a non-physical approach and wrote that they followed the methodology of Aalto et al. (2013). However, the title of this paper is "Spatial interpolation of monthly climate data for Finland: comparing the performance of kriging and generalized additive models". In fact, Aalto et al. tested the following methods: kriging with external drift (KED), generalized additive models (GAM), and GAM combined with residual kriging (GK). So, please, explain which method was, in fact, used in this paper. In addition, it is important to explain why another method was used to downscale wind speed and which were the "other variables" (page 7, line 1) used in this procedure?

Firstly, we decided to use the ERA-Interim data only for wind speed because of its much coarser resolution (0.75°×0.75°) compared to the Finnish observational data set (0.1°×0.2°). Unfortunately, wind speed observations suffer both spatial and temporal inhomogeneity for which reason producing a high-resolution observational gridded wind speed data set cannot be recommended. Moving completely to the resolution of the ERA-Interim data set would only dramatically reduce the resolution of our data, as we now have approximately 28 grid cells within each ERA-Interim grid cell.

Secondly, daily mean values were used because climate model outputs are typically provided as daily and monthly mean values, i.e., afternoon values are usually not available. Note, that we, however, used daily maximum temperature as a proxy for afternoon temperature. Actually, we already had to abandon one model because relative humidity for that model was available only on the monthly scale.

The historical conflagrations were studied independently and this part of the study can be easily removed. This extension was done because during previous decades, larger fires had occurred than were present in our data set. The recent Swedish fire was included because it has raised a question "whether this kind of conflagration could be possible in Finland nowadays or in the future?"

Perhaps the most important practical reason to use GCMs instead of RCMs was that at the time when the models were selected, outputs of RCMs based on the recent CMIP5 project were not yet widely available. Indeed, downscaling of the RCM outputs onto the Finnish grid would have been somewhat smoother than downscaling of the GCM outputs. On the other hand, the climate change signal in the RCM simulations is largely determined by the driving GCM. Moreover, RCM simulations tend to suffer severe wet bias in Europe; they often produce 30–50 % more precipitation than the driving GCM (see, e.g., Fig. 1 in Lehtonen et al. 2014a). As we chose the models based on their performance compared to the observed mean climate in order to reduce the need for bias correcting, this would have made it more difficult to choose the models with this approach. That is because choosing an RCM that well reproduces observational mean precipitation would have required choosing an RCM driven by a GCM with severe dry bias. To conclude, we think that both the use of RCMs or GCMs could be well reasoned here but they both also have their own drawbacks. In any case, in order to illustrate the model-based uncertainty that was one of our main goals in this study, it is important to use several models driven with different GCMs – or directly different GCMs like done here.

We chose 0.1°×0.2° (~10 km×10 km in Finland) grid to make the grid cells approximately as long both in the east–west and south–north directions instead by using, e.g., 0.1°×0.1° grid (~10 km×5 km in Finland).

We applied kriging with external drift (KEV) which was found to have in general the best performance in interpolating climate data over Finland among the methods tested by Aalto et al. (2013). We agree that we should be more specific here.

Another method was used in downscaling the wind speed values than other variables because in the case of other variables, point values were interpolated but reanalyzed wind speed values represent mean values over grid cells.

The other variables which were interpolated onto the 0.1°×0.2° grid with KEV were 2-m air temperature (daily mean, maximum and minimum), daily mean 2-m relative humidity and daily precipitation (page 4758, lines 23–24).

3. The confidence on the findings of this study is clearly affected by several factors. The most important of them are: (i) the short study period; (ii) the very small number of fire events; (iii) the development and robustness of the fire models (eq. 3 and eq. 4) based on a not very clear relationship with the fire indices. Eighteen years of data means just 18 data points to fit which could lead to erroneous conclusions, especially with so few number of large fires in Finland. In fact, 112 large fires (i.e., fire with $10 \leq BA<200$ ha), in the 1996 – 2014 means 6 large fires/year which is simply too low. In addition, the authors do not present a histogram of the large fires allowing the reader to understand how many of these fires are close to 10 or to 200 ha. Finally, I'm particularly concerned with the fire models' development and evaluation. Who intends to simulate a process with climatic variables has to describe the climate of the study area and unequivocally demonstrate the dependence of this process with these variables. The climate description at monthly scale is of fundamental importance to clearly understand the impacts of climate and climate change on fire activity in Finland and to understand the selection of DSR and MSR as predictors. For example, Fig 1 present the projected changes in the period of 7 months (April – October) which is a too long period and raised several questions/doubts. For example, the projected increase in precipitation and air temperature is identical in all months (between April and October) or is due to an increase in just a few months, eventually out of the fire season? The predictor selection process must be clarified. How were the predictors DSR and MSR selected? Did the authors test other potential predictors? It is quite intriguing that the selection of DSR and MSR as the predictors in an exponential model when, as recognized by the authors, "most of large forest fires are still ignited with relative low DSR". How was the DSR threshold of 15 determined? Did you test other thresholds? Which were the results? As the authors know, R is not the only or the best indicator of the quality of a model and not even an R equal to one ensures good performance of a model. It is important for the reader to know if the assumptions were tested. Differences between obtained values of rp and rs were not discussed/interpreted/validated. As the authors also know, Pearson's correlation (rp) is a statistical measure of the strength of a linear relationship while Spearman's correlation coefficient (rs) is a statistical measure of the strength of a monotonic relationship. Thus, it can be very dangerous to use rp to assess the strength of an exponential relationship. On the other hand, there are authors claiming that an rs equal to 0.39 reveals a weak association while an rs equal to 0.58 reveals a moderate correlation. This implies that the fire models (number of large forest fires and burned area) presented in this manuscript may not be as strong as the authors claim and, eventually, they should be used very carefully under current conditions and its use discouraged out of calibration range (future climate scenarios), specially without a previous proper assessment (e.g., with cross validation). The authors seems to be satisfied with R2 values of 0.67 and 0.81. However, these results mean that the models are only able to explain, respectively, 45 and 65% of the variance of number of large forest fires and burned area which must be addressed by the authors. Climate

change and its impacts are a matter of assessing the existence, magnitude and significance of changes between current and projected statistical distribution. Therefore it is of fundamental importance to assess the statistical significance of all the results obtained in this study, besides of the correlation coefficients. The values of the mean bias and RMSE may, or not, increase the confidence in these models as it much depends on other factors such as: the mean burnt area of the 112 large fires used in this study (which is unknown for the reader) and on the size of the time series. In this sense, what were the periods of calibration and validation of the models (please see the caption of Table 4)?

Our fire data set of 19 years includes a total of 19313 forest fires. We tested the relationship shown in Fig. 2 with different size of fires (all fires: N=19319; over 1 ha fires: N=1828; over 2 ha fires: N=895; over 5 ha fires: N=256; over 10 ha fires: N=112, over 20 ha fires: N=46, and over 50 ha fires: N=14). When all fires were included, the relationship between DSR and fire occurrence appeared to be nearly linear and the larger fires were excluded, the stronger became the power relationship (note that DSR itself has a power-law relationship with FWI). Moreover, when the sample size was increased by including smaller and smaller fires, the fire probabilities with a certain DSR value became much smaller in northern than southern Finland when the country was divided into sub-regions (we tested this with 2 and 4 sub-regions). With 10 ha threshold, we did not see it meaningful to reduce the sample size anymore by dividing Finland into two or more sub-regions because the relationship proved to be very similar both in northern and southern Finland if the country was divided into two sub-regions (the difference in the estimated fire probability being less than 15 % at the most). To conclude, we considered 10 ha as a good threshold because with this threshold there were not anymore discernible differences in the fire occurrence with a certain DSR value between different regions of Finland (unlike with smaller fires) and the shape of the relationship between DSR and fire occurrence was confirmed by studying the relationship with different size of fires which revealed that the larger fires are considered, the stronger the power-law relationship becomes, although the exact fit has larger uncertainty due to a smaller sample size. In addition to the power relation, we also tested fitting the linear relation and the exponential relation in the Eqs. 3 and 4. The power relation appeared to fit the data better than the other two relations.

About 45 % of all forest fires larger than 10 ha in the data set were larger than 20 ha as can be seen from Table 5.

It is true that projected changes on monthly time scale in climate variables shown in Fig. 1 do not precisely correspond to the mean change of fire season. Like mentioned in the text, precipitation may even decrease in the southern and eastern parts of the study region during midsummer months. In general, precipitation is projected to increase more in spring and autumn than in summer and this could be highlighted. For air temperature, somewhat smaller changes are similarly projected for midsummer.

Basically, we chose the FWI system because it has been widely implemented around the world and in many kind of environments. Instead of the FWI value itself, we used DSR because it is generally *acknowledged to reflect better the actual fire activity because it emphasizes the higher FWI values through the relation. This was also demonstrated in this study as the correlation between SSR and annual burned area was higher (~0.75) than reported by Venäläinen et al. (2014) between annual mean FWI and burned area (~0.60) in Finland. We also studied how well the observed fire activity on a seasonal scale correlated with different moisture codes of the FWI system and found out that duff moisture code and drought code were almost as good predictors for the fire activity than the SSR itself. Fine fuel moisture code had less predictability on a seasonal scale, and mainly for the number of forest fires, not for burned area. Probably on daily scale fine fuel moisture code is a much better predictor but we still wanted to prefer the FWI rating (and its power-law form, DSR) because it combines the effect of all other codes in the FWI system.*

We do not see contradiction between the facts that most of large fires are ignited with relatively small DSR values even though the probability of ignition increases exponentially with increasing DSR. That is, because also the occurrence of days with high DSR value decreases exponentially with increasing DSR and even much more rapidly than the total number of large forest fires. As a result, the share of days with large fires indeed increases with increasing DSR although most of large fires occur with relatively low DSR. It was a subjective choice to choose the threshold 15, but this threshold was chosen because there had occurred only 2 early season large fires and 7 late season large fires with a DSR value of 15 or higher. So, only 8% of all large fires fell into this category but more importantly, only 0.2% of considered days. As it can be seen from Fig. 2, the relationship between large fire activity and DSR becomes less clear due to a decreasing sample size already when DSR exceeds about 10, but still 1.2% of days expressed a DSR value of 10 or higher. However, with smaller fires included, the relationship follows more closely the power relation, with all fires included up to a DSR about 20. In any case, we considered DSR values over 15 as marginally important as they represent only 0.2% of days. This was confirmed by repeating the analyses by extrapolating the relationship of Eq. 3 with DSR values above 15 which showed that the overall results were not much affected; only on the driest future years the number of large fires would have been increased even more than 50% if the power relation would have been extrapolated. Typically, the increase in the number of large fires would have been about 10% within the reference period and 10–30% during the future periods.

It is clear that climate variations can explain only a part of variations in fire activity. If climate variations explain roughly about 50% of variations in fire activity, as it seems to be the case here, it is clear that a model estimating the impact of climate change on fire activity can not explain more than a half of the future fire activity. Certainly this issue could be emphasized more clearly. Moreover, the reviewer is right that it is questionable whether similar relationship than nowadays between MSR and burned area would still hold with MSRs much higher than observed by this time.

The calibration period of the models was 1996–2014.

4. The manuscript is not always well written. There are several unnecessary repetitions (e.g., Caption of Table 4) and citations. The authors use and abuse of relative and subjective concepts (e.g., strong, majority, most, a few) without clear/objective definition which is strongly

discouraged. There are several sentences as well as table captions beginning with numerals which, in my opinion, should be avoided. The fire related concepts are not the ones usually used within the fire community. For example, in the title, the authors claim they are studying fire risk when, in fact, they are studying fire danger as recognized in page 4, lines 24 – 26. Why use fire source instead of fire cause? I suggest the use of the Glossary of Wildland Fire Terminology.

We agree with the reviewer that the text can be improved, e.g., by reducing the use of subjective words.

Some more specific comments

While the problems identified above are not corrected it does not seem appropriate to discuss detailed issues. I include just a few comments/questions to illustrate/underline some of the subjects discussed before.

1. Page 4, lines 24 – 26: "Hence, there exists a clear need to update the fire danger assessment by using several models instead of the usually applied multi-model mean approach." An explanation on how the changes presented in the manuscript were computed is advised. Please note the caption of Figure 1 "Dots indicate the multimodel mean change . . . ";

Please note that we also show the range of projected changes among the selected models. So the uncertainty ranges are not derived, e.g., from a single model or by assuming the climate variability to remain constant.

2. Page 6, lines 2 – 3: Information about the size and location of the Åland Islands in relation to mainland Finland should be provided for non-Finnish readers; it is also important (Page 12, lines 2 – 7) to include the latitude and longitude in Figure 3 and 7;

Thank you for this comment, we agree that the location of the Åland Islands should be mentioned.

3. Page 6, lines 6 – 8: "Those fires which exact coordinates were not reported were located in the middle of the municipalities where the fires reportedly had occurred". Does this mean that, in the case where the spatial coordinates of the fire were missing, it was assumed to be equal to the coordinates of the centroids of the municipality where the fire had initiated?

Yes.

4. Page 6, lines 8 – 9: Which "other types of wildland fires" besides forest fires were excluded? What is your definition of forest fire? Please clarify it.

Other types of wildland fires include, for example, peat pile fires, grassland fires, bulrushes fires, small fires on parks in cities and towns etc. Especially grassland fires are common in early spring before vegetation has developed and before the onset of the forest fire season. We picked up those *fires that were reported as "forest fires" in the database, i.e., they reportedly occurred in forested environment.*

5. Page 8, lines 10 – 18 and Figure 1: Why present projected changes in April-October daily maximum temperature, mean relative humidity and mean wind speed? It will be much more informative to present the projected changes of air temperature, relative humidity and wind at noon because these are the variables used to compute the Canadian fire weather indices. Please define "fairly", "robust" and "coherent" and explain why projections for relative humidity are "fairly robust" and projection for other variables are not. Please explain why projections of wind speed are not coherent?

Unfortunately, climate model data is provided as daily mean values. For temperature, also daily maximum and minimum values are available and we thus use daily maximum temperatures in our analyses. We note that the projected change in daily maximum temperature does not deviate much of that for daily mean temperature. For wind speed, part of models project decrease and some models project increase while for relative humidity, the models tend to project more decrease than increase and for temperature and precipitation all models project coherently increase. To conclude, we think that a projection for a certain variable is coherent if all models unanimously project either decrease or increase but the projection is not coherent if about half of the models project decrease and rest of the models increase.

6. Page 11, lines 16 – 18 and Table 3: The coefficients of eq. 4 were defined/used (please see the manuscript and table caption) or estimated? In the second case, please explain how;

The coefficients were defined by fitting a power relation to the data set, similarly as for the Eq. 3.

7. Page 11, lines 16 – 18 and Table 3: The number of decimal places of coefficients a and b seems to be exaggerated. How many decimal places of the coefficients a and b are significant figures?

I would not expect that extra decimals cause any harm. For example, for July with MSR=4, the estimated burned area increases by about 3% if only one decimal is used in the coefficients instead of two decimals, so I would not say that using two decimals is too much.

8. Page 12, lines 22 – 25: What do you mean by "rather similar"? With such a small number of large fires at national level is it advisable to consider sub-regions? How many sub regions were considered and how many large fires occur in each sub-region?

We tested this with two and four sub-regions. We agree that four sub-regions would be clearly too much for such a small number of large fires. With two sub-regions, the fire probabilities did not diverge more than about 15% at the maximum between the two sub-regions with DSRs below 15. For smaller fires, it was clear that the fire probabilities with a specific DSR values were higher in more populated sub-regions than in less populated regions. This was one reason, why we chose to study the fires larger than 10 ha in more detail.

9. Page 14, lines 3: The value of 1.5% written in the manuscript, is not shown in Table 5. I believe it results from the sum of 0.7 with 0.8 which corresponds to fires with BA >=10 ha. This should be clearly explained in the manuscript;

Yes, your assumption is right.

10. Page 15, lines 8 – 9: I believe this sentence should be changed because it is not in accordance to the sentences in the following lines;

We believe that the referee is referring to the following two sentences: "Regionally, the forest fire danger is projected to increase rather similarly throughout Finland (Fig. 7). Under the RCP8.5 scenario, multi-model mean SSR averaged over April–October period increases in the south from about 2–3 to 4–6 and in the north from about 1 to 2 until the end of the 21st century." We disagree with the reviewer, because the change is in this case about 100% both in the south and north. although this could be explicitly stated.

Technical corrections

At this stage, it is not opportune to suggest any technical correction.