

Reviewer 1 – comments:

This manuscript addresses the risk of late spring frost damage to the leaves of temperate trees. The authors use long-term leaf-out phenology records of six European tree species to study the environmental factors that affect the severity of late spring freezes and how climate change is altering these patterns. The main finding is that spring temperature and distance to the coast are the main drivers of spring frost risk, but climate warming is currently altering these trends, complicating forecasts of future spring frost damage. In addition, early-leafing species are found to be more prone to spring freezes with warming, whereas no change or a decrease in spring frost risk was found for late-leafing species. The manuscript addresses an interesting, highly relevant question to better understand the consequences of climate change for plant growth.

We thank the reviewer for this positive feedback and for helping us improve the manuscript. Based on this review we have updated the methods and definitions for all of the models. We hope our current manuscript better addresses our questions and improves our understanding of the effects of climate change on plant growth.

Yet, I have two major concerns with respect to the methodology of the manuscript. First, the relevant period during which the leaves of trees are susceptible to freezing events is defined as the 12 days before leaf unfolding (BBCH11). This makes no sense to me. I agree with the assumption that young leaves are more susceptible to freeze events than older leaves and the authors correctly state that “plants are most susceptible to damage from freezing temperatures between budburst and full leafout”. However, BBCH 11 is NOT full leafout, it is the stage where the first visible leaf stalk is visible, but not on all buds, only on the first opening buds. Plants are definitely still susceptible to frost after that event. In fact, they should even be more vulnerable to frosts after BBCH11 was reached because the number of buds opening will increase drastically only after BBCH11. The next phenophase available in the PEP725 database (BBCH 13) is the stage where 50% of leaves are unfolded, which means that even at this later stage, 50% of leaves are not yet unfolded. So, to me the only way of determining whether the results are robust is to use another approach of calculating the relevant freezing period. Spring freezes after BBCH11 should definitely be accounted for.

We appreciate the concern of Reviewer 1 and have worked to address it by modifying our methodology to include the 12 days before BBCH 11 through 12 days after BBCH 11 as the period of risk. We agree that BBCH 13 is more appropriate—though still not full leafout—but there were only 12,417 observations for BBCH 13 compared to 755,087 observations for BBCH 11 so we decided to continue using BBCH 11 as an indicator for individual phenology.

Second, the study assumes that spring freeze damage will occur at -2.2 C for all species. Yet, leaf-killing temperatures are species-specific (Lenz et al., 2016; Muffler et al., 2016), which should be accounted for when testing for differences among species (Zohner et al., 2020b). Leaves of early-leafing species are usually more freezing resistant than those of late leafers (Muffler et al., 2016). I can see that this problem was partly solved by additionally using -5C as freeze threshold, but this additional analysis only changes the general freezing threshold and still doesn’t account for differences among species. The leaves of Aesculus, Alnus, and Betula are more resistant to frost than the leaves of Fagus, Fraxinus, and Quercus (data in Muffler et al., 2016; Zohner et al., 2020a, not yet published but attached to this review). It would therefore be interesting to see whether the species-level results still hold when using a freezing threshold of -2.2 C for the latter group and -5C for the first group of species.

We thank the reviewer for this insight and new idea to improve the manuscript. We have included a new model that uses varying thresholds for early- versus late-leafout species. We have included a new paragraph to the Results section *Sensitivity of results to duration of risk and temperature thresholds* (lines XX-XX) to address these results:

Results, again, remained consistent in our varying threshold model with all predictors contributing to risk: mean spring temperature (-10.38% for every 2° or -0.65 ± 0.13 probability of risk/standard unit), distance from the coast (2.41% for every 2° or 0.18 ± 0.14 probability of risk/standard unit), elevation (7.48% for every 2° or 0.64 ± 0.14 probability of risk/standard unit) and NAO (3.74% for every 2° or 0.28 ± 0.12 probability of risk/standard unit). There was a slight increase in false spring risk due to the residual effect of climate change across all six species (29.69% increase or 1.1832854 ± 0.06 probability of risk/standard unit; Figure S6 and Table S9). The late-leafout species (i.e., *Fagus sylvatica*, *Quercus robur*, *Fraxinus excelsior*) experienced more false springs than the early-leafout species (i.e., *Aesculus hippocastanum*, *Alnus glutinosa*, *Betula pendula*), though after climate change all species experienced a more similar magnitude of risk (Figure S7).

l. 53: reference needed for “but budburst has advanced roughly twice as fast”. You are saying budburst advanced by, on average, 5 days per decade. From my own observations it’s more like 23 days per decade, which would then be similar to the last spring freeze advances.

We thank the reviewer for finding this error and have updated the sentence and have added citations (lines XX-XX):

In Germany, for example, the last freeze date has advanced by 2.6 days per decade since 1955 (Zohner *et al.*, 2016), but budburst has advanced 4.3 days per decade in Central Europe (Fu *et al.*, 2014; Vitasse *et al.*, 2018).

l. 360–368: I don’t get the point of this discussion on species ranges. The phenological data used here came mostly from Germany and didn’t reflect species ranges at all.

We appreciate the reviewer’s concern and have updated the language in the Discussion section *Variation in risk across species* (lines XX-XX) to provide additional clarity. We hope that this modified section is easier to follow and understand:

Though our study focuses on Central Europe, overall habitat preference and range differences among the species could also explain some of the species-specific variation in the results (Chuine *et al.*, 2001), but would require data on more species—and species that vary strongly in their climatic and geographic ranges—for robust analyses. The ranges of the predictors are similar across species within our dataset, but *Betula pendula* extends to the highest elevation and latitude and spans the greatest range of distances from the coast (Figure 2), while *Quercus robur* experiences the greatest range of mean spring temperatures. Within our species, *Betula pendula* has the largest global distribution, extending the furthest north and east into Asia. The distribution of *Fraxinus excelsior* extends the furthest south (into the northern region of Iran). These global range differences could potentially underlie the unexplained effect of climate change seen

in our results and why the climatic and geographic factors failed to explain all of the variation in false spring risk for our species. Future research that captures these spatial, temporal and climatic differences across myriad species could greatly enhance predictions and help us understand these residual effects of climate change. Such research may be particularly useful if it connects how range and habitat differences translate into differences in physiological tolerances and the underlying controllers of budburst and leafout phenology—the factors that proximately shape false spring risk.

Reviewer 2 – comments:

The authors describe a statistical analysis on the probability of false springs for an extensive set of drivers, in this respect scaling beyond previous studies. The authors conclude that false spring risks are strongly associated with mean spring temperature and distance from coastal waters, with increased risks for early leaf-out species.

I've appreciated the rather technical write-up, and in particular the future perspectives which were well defined and outline new research avenues. However, I would like to see some things clarified.

We thank the reviewer for this appreciation and for helping us clarify certain sections and improve our figures. Based on this review we have included data availability, updated the methods and definitions for all of the models and enhanced our figures to be more colorblind-friendly.

First and foremost a practical issue. Given the use of open data, open software I will insist on the open nature of the analysis executed and all source code to be shared. Given PEP725 redistribution rules data can't be distributed, however code to compile the data in the correct data format can. I suggest to use a repository such as github / gitlab / bitbucket to store your code (and revisions) in their native format (i.e. no copied code in txt files, but .R files). All this should ensure reproducibility, transparency and facilitate re-use of code within a research or educational setting.

We completely agree with the reviewer about this concern and apologize for failing to include our data in the first submission. We have included a new section *Data, Code & Model Output*: (lines XX-XX) with this information:

Raw data will be available via KNB upon publication and are available to all reviewers upon request. Raw data, Stan model code and output are available on github at <https://github.com/cchambe12/regionalrisk> and provided upon request.

Furthermore, to what degree do you think the use of leafout instead of budburst influences the analysis? In particular, the data does not take into account delays in development / leafout due to prior false spring damage (e.g. Gu et al. 2008 / Hufkens et al. 2012 / Augspurger 2007). It could therefore be argued that a fixed offset could influence the results. Although acknowledged on lines 272 onward, using the varying offsets, the use of these undamaged developmental responses is a missing confounding factor which should be discussed.

We appreciate the reviewer's concern and have modified our methodology to include the 12 days before BBCH 11 through 12 days after BBCH 11 as the period of risk, which we hope

will capture this possible delay in development due to prior false spring damage.

Also, on line 140 - 145, it is mentioned E-OBS data is used but I see no mention of lapse rate corrections for altitude to correct for site by site differences (e.g. Basler 2016)? Given the absolute temperature thresholds I think this applies.

We thank the reviewer for this attention to detail and for pointing this out. We have included more information about our temperature data and a citation in Methods section *Climate Data* (lines XX-XX):

E-OBS version 16 incorporates station altitude in the interpolation scheme, thus spatially explicit information on day-to-day variability in the environmental lapse rate is captured (Cornes *et al.*, 2018).

Finally, some smaller issues:

I would increase readability of the figures by avoiding pastel colours. Although I'm not color-blind I'm not sure how well this would work for those who are.

- Use varying line styles instead of colour if possible (Figure 5)
- Don't use colours where not needed (Figure 2 and 3 - species labels)
- Make sure that points are properly scaled (smaller for Figure 2), or better regrid the data to show spatial trends more consistently
- Increase the font size on all figures (for axis labels etc)

I think all of these issues are easily addressed, either in discussion (given data limitations) or with the necessary model / driver data adjustments. Good luck, stay safe !

We appreciate the reviewer's concerns about our figures and have updated the colors using the viridis package in R, which is colorblind-friendly. Additionally, we have removed unnecessary colors, rescaled the points in Figure 2 and increased the font size on all figures.

References

- Chuine I, Aitken SN, Ying CC (2001) Temperature thresholds of shoot elongation in provenances of *Pinus contorta*. *Canadian Journal of Forest Research*, **31**, 1444–1455. doi: 10.1139/x01-072. URL <http://dx.doi.org/10.1139/x01-072>.
- Cornes RC, van der Schrier G, van den Besselaar EJM, Jones PD (2018) An Ensemble Version of the E-OBS Temperature and Precipitation Data Sets. *Journal of Geophysical Research: Atmospheres*, **123**, 9391–9409. doi:10.1029/2017JD028200. URL <https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2017JD028200>.
- Fu YH, Piao S, Op de Beeck M, *et al.* (2014) Recent spring phenology shifts in western central europe based on multiscale observations. *Global Ecology and Biogeography*, **23**, 1255–1263. doi:10.1111/geb.12210. URL <http://dx.doi.org/10.1111/geb.12210>.
- Vitasse Y, Schneider L, Rixen C, Christen D, Rebetez M (2018) Increase in the risk of exposure of forest and fruit trees to spring frosts at higher elevations in Switzerland over the last four decades. *Agricultural and Forest Meteorology*, **248**, 60 – 69. doi:<https://doi.org/10.1016/j.agrformet.2017.09.005>.
- Zohner CM, Benito BM, Svenning JC, Renner SS (2016) Day length unlikely to constrain climate-driven shifts in leaf-out times of northern woody plants. *Nature Climate Change*, **6**, 1120–1123. doi:10.1038/nclimate3138.