

Reviewer 1 – comments:

The topic of false springs warrants additional attention, such as provided by this manuscript. I think the manuscript provides a useful contribution by summarizing available information from the literature as well as some original results. The bibliography is good, the figures are useful, and I think relevant literature has been discussed. I have made a lot of editorial suggestions on the PDF, and indicated a few points for the authors to consider.

We appreciate the concerns of Reviewer 1 and have worked to address them. We have inputted all corrections from the PDF including grammatical errors, improving sentence structure, removing the noted split infinitives and improving the colors for Figure 2. In regards to addressing tissue types we added the suggested reference and worked on structuring the text to stress the importance of focusing on vegetative parts (lines 65-74).

The flowering and fruiting phenophases are generally more sensitive to freezing temperatures than vegetative phases (Inouye, 2000; Augspurger, 2009; CaraDonna, Paul J and Bain, Justin A, 2016; Lenz *et al.*, 2013), but false spring events that occur during the vegetative growth phenophases may impose the greatest freezing threat to deciduous plant species. It can take 16-38 days for trees to refoliate after a spring freeze (Gu *et al.*, 2008; Augspurger, 2009, 2013; Menzel *et al.*, 2015), which can detrimentally affect crucial processes such as carbon uptake and nutrient cycling (Hufkens *et al.*, 2012; Richardson *et al.*, 2013; Klosterman *et al.*, 2018). Additionally, plants will suffer greater long-term effects from the loss of photosynthetic tissue, which could impact multiple years of growth, reproduction, and canopy development (Vitasse *et al.*, 2014; Xie *et al.*, 2015). For this reason, we will focus primarily on spring freeze risk for the vegetative phases, specifically between budburst and leafout, when vegetative tissues are most at risk of damage.

We added text to better explain specific morphological strategies (i.e., the use of trichomes in young leaves). We also appreciate the comment about *Syringa vulgaris* not being a native plant and agree this is a crucial point to mention. We added text (lines 107-109) in the *Measuring False Spring: An example in one temperate plant community* section and again in a new section *The Future of False Spring Research* (lines 256-257):

While the SI-x data (based on observations of early-active shrub species, including lilac, *Syringa vulgaris*) may best capture understory dynamics, the PhenoCam and observational FSI data integrate over larger canopy species and includes native species — unlike SI-x.

For studies looking to understand community shifts in species range shifts, it is important to recognize the limitations of using SI-x. The SI-x metric uses non-native, small shrub species.

Reviewer 2 comments:

Chamberlain et al. presented a perspective paper on complexities and challenges facing the community to study the risk of spring frost, termed as false spring risk. While the topic is certainly very interesting to a broad readership, I have several major concerns on definitions and reasoning made in the manuscript. Therefore, I cannot recommend publishing the manuscript at its present form. First of all, the damaging late spring freeze not only affects

early leaves but also flowers developing during the frost period (Augspurger, 2009). The focus on leaves is ok, but it has to be clarified from the very beginning to the conclusions that the manuscript address only the leaf part of spring frost damage.

We appreciate Reviewer 2's concerns and agree that our focus on vegetative phases was not well constructed in its original form and restructured the document and bolstered the argument to focus on vegetative phases (lines 65-74) as afore mentioned with Reviwer 1's comments:

The flowering and fruiting phenophases are generally more sensitive to freezing temperatures than vegetative phases (Inouye, 2000; Augspurger, 2009; CaraDonna, Paul J and Bain, Justin A, 2016; Lenz *et al.*, 2013), but false spring events that occur during the vegetative growth phenophases may impose the greatest freezing threat to deciduous plant species. It can take 16-38 days for trees to refoliate after a spring freeze (Gu *et al.*, 2008; Augspurger, 2009, 2013; Menzel *et al.*, 2015), which can detrimentally affect crucial processes such as carbon uplake and nutrient cycling (Hufkens *et al.*, 2012; Richardson *et al.*, 2013; Klosterman *et al.*, 2018). Additionally, plants will suffer greater long-term effects from the loss of photosynthetic tissue, which could impact multiple years of growth, reproduction, and canopy development (Vitasse *et al.*, 2014; Xie *et al.*, 2015). For this reason, we will focus primarily on spring freeze risk for the vegetative phases.

The structure of the manuscript needs a careful reconstruction. Readers cannot well deduce the reasoning of the manuscript at its present form. For example, the sub-titles of the sections are poorly made. One cannot deduce the logic flow by putting them together, which arose much questioning on the structure: Why measuring false spring comes before defining false spring? Should definition comes first? I have to go through the manuscript for quite a few time in order to figure out the logic links between different sections. Still, it keeps me wondering why the authors organizing the manuscript in its current way. The manuscript has developed arguments on three key parameters: the date of budburst, the date of full leafout, the temperature threshold of frost damage. However, rather than bringing up all the three parameters, the issue of full leafout date only emerges after partial discussions of plant strategy to avoid frost risk. This organization is confusing. Why should the definition on the duration of sensitive period comes after the plant physiology strategy to avoid frost risk. Should shortening of the sensitive period also a physiological strategy to avoid frost, just like the sensitivity of budburst to photoperiod?

We agree with the reviewer's concerns regarding overall structure of the manuscript. We have addressed this by changing the names of headings and sub-headings and by also moving around sections per his/her recommendation. The original organization was not as smooth as it could be. In regards to the order of *Defining Vegetative Risk* after the section on plant tolerance and avoidance strategies, we have now made *Defining Vegetative Risk* a sub-heading and added text to explain that it could also be an avoidance strategy against false springs (lines 165-168):

With spring advancing, species that have shorter durations of vegetative risk may avoid false springs more successfully than species that have much longer durations of vegetative risk. Understanding the various physiological and phenological mechanisms across species are crucial for species- or site-specific studies as well as ecosystem-wide models.

There are several issues related to Figure 1. How budburst dates are derived from the SI-x, Observational and Phenocam data are missing. It is well known that the budburst date is very sensitive to detecting methods and data sources. It should contain key information on the date retrieval in order for readers to determine whether they are comparable. The references cited here are not peer-reviewed, and they have neither weblinks nor doi. There could be an obvious potential issue on comparing them, where the ground observation are individual tree-based, while Phenocam is canopy/community based. Should they be treated the same?

We thank the reviewer for his/her concerns. The aim of this section was to stress that these three methods are not comparable and encourage researchers to carefully choose the best method for their study. We have added more information to the references and have added text to hopefully stress these issues (lines 115-129):

Yet, in contrast to our three metrics of spring onset for one site, most FSI work currently ignores variation across functional groups — instead using one metric (often from SI-x data) of spring onset and assuming it applies to the whole community of plants (Allstadt *et al.*, 2015; Marino *et al.*, 2011; Mehdipoor & Zurita-Milla, 2017; Peterson & Abatzoglou, 2014). The risk of a false spring varies across habitats and with species composition since spring onset is not consistent across functional groups (Martin *et al.*, 2010). Therefore, one spring onset date cannot be used as an effective proxy for all species and the three methodologies are not interchangeable.

Rather than use any metric to evaluate spring onset, we encourage researchers to first assess the forest demographics and functional groups relevant to the study question. From there, researchers can choose the most appropriate method to estimate the date of budburst to determine if a false spring could have occurred. However, as we outline below, considering different functional groups is unlikely to be enough for robust predictions in regards to level of damage from a false spring, especially when trying to determine how frosts are shaping the life history of plants. In such cases, it will also be important to integrate species differences within functional groups and to consider the various interspecific tolerance and avoidance strategies that species have evolved against false springs.

The scale issue can also be found in later parts of the manuscript. For example, Figure 5 clearly shows that studies at regional scale may suffer from how to define the sensitive period at regional scale. How should this question be addressed? Discussions and explorations should help improve the manuscript for readers' benefits.

This is a key issue and we agree with the reviewer that further discussion is necessary. We have added more on this issue throughout and have added a new section *The Future of False Spring Research* to address this more directly (lines 227-249):

False springs are one of the strongest driving factors for species range limits (Sakai & Larcher, 1987), thus are crucial to implement in most — if not all — temperate plant studies. The FSI, single-metric approach is better than not including spring frost risk at all and would be a great place to start for most ecosystem models. In such models, the SI-x would provide researchers with predicted shifts in frequency of false springs under emission scenarios. The Ecosystem Demography (ED) model already integrates phenology data by functional group (Kim *et al.*, 2015; Moorcroft *et al.*, 2001), by adding last freeze date information, FSI could then be evaluated to predict false spring occurrence with predicted shifts in

climate. By including some sort of proxy for false spring risk, the ED model, and similar models, could better inform predicted range shifts.

For studies looking to understand community shifts in species range shifts, it is important to recognize the limitations of using SI-x. The SI-x metric uses non-native, small shrub species. And by using only FSI, there is no measure for actual damage, it is simply an estimation of potential damage. PhenoCam data is a way to bridge the gap between ground observations and SI-x data. It is possible for researchers to look through every image or to simply use a specific greenness parameter to identify greenup (Richardson, 2018). By using the greenness parameter with FSI, researchers can better approximate community-specific damage.

But, sometimes using SI-x and even FSI will not work at all, such as for studies that are looking to understand life-history theory. PhenoCam images can also capture the shift in greenness: initial greenup, defoliation due to frost or herbivory, then refoliation (Richardson *et al.*, 2018). However, using observational data is likely the best method to capture different functional types and lifestages. Through observational data, researchers can capture the variations in temperature thresholds, species- and lifestage-specific tolerance and avoidance strategies and climatic effects. False spring events could have large scale consequences on forest recruitment, potentially impacting juvenile growth and forest diversity. By using on-the-ground observations, researchers can more accurately measure the level of damage and the variation across lifestages.

Figure 5 use a climate dataset with its source not clearly presented. Is it from a gridded dataset or climate stations? If it is from a gridded dataset, which dataset is used? What are the spatial-temporal resolution and temporal coverage of it?

We agree with the reviewer’s concerns and have now added more information about the methods and data we used for this portion of the manuscript (lines 196-203) as well as reference the supplement in the caption:

We analyzed five archetypal regions across North America and Europe. We assessed phenology data from USA-NPN SI-x data and phenological studies that monitored budburst and leafout across Europe. We then collected climate data by downloading Daily Summary climate datasets from the NOAA Climate Data Online tool (NOAA, 2017) and calculated the number of years that fell below -2.2°C within the budburst to leafout date range for each region. We found that some regions experienced harsher winters and greater temperature variability throughout the year (Figure 5 e.g., Maine, USA), and these more variable regions often have a much higher risk of false spring than others (Figure 5 e.g., Lyon, France).

It is not clear why winter chilling requirement is a strategy to avoid spring frost. Is warmer winter correlated with warmer spring?

We appreciate the concern from the reviewer regarding chilling and false springs and attempted to add more text to clarify (lines 146-149):

Warm temperatures too early in the winter (i.e. in February, or even January in the Mediterranean) will not result in early budburst due to insufficient chilling (Basler & Körner, 2012), thus reducing the risk of false spring damage.

Line 36: this research -> these researches.

We agree there are issues with this sentence. We have now improved it by removing the redundancies.

To produce accurate predictions, however, researchers need methods that properly evaluate the effects of false springs across diverse species and climate regimes.

Reviewer 3 comments:

This is an interesting opinion piece which makes some useful points about assessing the risk of "false springs" from climate data. In general I think the key points made here are important, and the piece is quite well written - the main take home point is that the factors influencing false spring risk are fairly complex and include species identity, life stage, varying phenological cues as well as regional climate variability. As such, I think that with some revision, this paper is worth publication in GCB. However, although the paper is strong on critique, I think it could be more constructive in terms of suggestions for directions of future research, and the structure could be improved. A good start would be to review why we need to define false spring risk in the first place, and why we might need a "false spring index" that can be applied across ecosystems (if indeed we do). The issues identified within the section titled "Defining False Spring" are largely due to the difficulties in defining a single date for "spring" across a whole ecosystem. As the authors recognise, different species come into budburst and leaf out at different times within the same ecosystem, often following different cues; there is also variation within species. As a consequence, no two methods for measuring the onset of spring (including field-based and satellite measures) are likely to produce the same date. the question is why do we need a single ecosystem-wide index of false spring, and is there any meaningful interpretation of such an index? This issue isn't really tackled, which is a shame.

We appreciate the reviewer's comments on our lack of direction for future studies and we agree, we failed to clearly state our suggestions. We have added a paragraph at the end of section *Measuring False Spring: An example in one temperate plant community* (lines 120-126):

Rather than use a single metric (i.e., FSI), false spring studies should first assess the forest demographics and functional groups relevant to the study question. From there, researchers can choose the most appropriate method to estimate the date of spring onset to determine if a false spring could have occurred. However, as we outline below, considering different functional groups is unlikely to be enough for robust predictions in regards to level of damage from a false spring. It will also be important to integrate species differences within functional groups and to consider the various interspecific tolerance and avoidance strategies that species have evolved against false springs.

The authors also point out that flowering and fruiting are generally more sensitive to false spring events than vegetative phases, but all examples mentioned in the paper are the vegetative phase.

All reviewers mentioned this flaw in the manuscript. We have addressed this issue by adding more to our argument to focus on vegetative phenophases and have moved this portion

earlier in the text. It is now in section *Defining False Spring*.

The problem identified is a common one in ecology - on the one hand there is the modeller and geographer's desire to create an index that has a general application, that can be mapped and modelled across ecosystems and applied to any region and time period, but on the other hand the natural instinct of field ecologists is to point out that things are inevitably more complex, and each species (and life-stage) within an ecosystem will respond to the environment in an individual manner. What I think is missing from this opinion piece is where the author's opinion is the balance between these viewpoints. Are ecosystem-wide false spring indices useful? If so, how should they be constructed to maintain generality while incorporating the characteristics of many species? Where do the opportunities for constructing more general models lie? Where are our knowledge gaps?

We really appreciate the reviewer's comments/concerns in regards to our suggestions, especially in regards to different types of studies. We have added more throughout the document to clarify these issues and our suggestions and have also added an entirely new section *The Future of False Spring Research* (lines 227-249) to address these concerns more directly, which was also a concern of Reviewer 2's:

False springs are one of the strongest driving factors for species range limits (Sakai & Larcher, 1987), thus are crucial to implement in most — if not all — temperate plant studies. The FSI, single-metric approach is better than not including spring frost risk at all and would be a great place to start for most ecosystem models. In such models, the SI-x would provide researchers with predicted shifts in frequency of false springs under emission scenarios. The Ecosystem Demography (ED) model already integrates phenology data by functional group (Kim *et al.*, 2015; Moorcroft *et al.*, 2001), by adding last freeze date information, FSI could then be evaluated to predict false spring occurrence with predicted shifts in climate. By including some sort of proxy for false spring risk, the ED model, and similar models, could better inform predicted range shifts.

For studies looking to understand community shifts in species range shifts, it is important to recognize the limitations of using SI-x. The SI-x metric uses non-native, small shrub species. And by using only FSI, there is no measure for actual damage, it is simply an estimation of potential damage. PhenoCam data is a way to bridge the gap between ground observations and SI-x data. It is possible for researchers to look through every image or to simply use a specific greenness parameter to identify greenup (Richardson, 2018). By using the greenness parameter with FSI, researchers can better approximate community-specific damage.

But, sometimes using SI-x and even FSI will not work at all, such as for studies that are looking to understand life-history theory. PhenoCam images can also capture the shift in greenness: initial greenup, defoliation due to frost or herbivory, then refoliation (Richardson *et al.*, 2018). However, using observational data is likely the best method to capture different functional types and lifestages. Through observational data, researchers can capture the variations in temperature thresholds, species- and lifestage-specific tolerance and avoidance strategies and climatic effects. False spring events could have large scale consequences on forest recruitment, potentially impacting juvenile growth and forest diversity. By using on-the-ground observations, researchers can more accurately measure the level of damage and the variation across lifestages.

Neither the conclusion nor the abstract seem to capture the main argument of the paper to me - for example, the abstract states that "we highlight how species, life stage and habitat differences contribute to the damage potential of false springs", but the body of the paper actually says rather little about life stage or habitat differences. Likewise, the first line of the conclusion states that "temperate forest trees are most at risk to frost damage in the spring due to the stochasticity of spring freezes" - this may well be true, but it is not a key conclusion of the body of the paper at present. I would recommend that the authors rewrite sections of this opinion paper to make it clear what their opinion is here - what direction should future research take, and what is the ultimate aim?

We agree with the reviewer that the Abstract and Conclusion fail to accurately portray the essence of the paper. We have updated the language, removed the line "temperate forest trees are most at risk to frost damage in the spring due to the stochasticity of spring freezes", and have added more suggestions to the conclusion.

Minor comments:

The examples given in the figures are great, but sometimes lack detail on methods - for example, in figure 5, which species do the observational studies come from? Given the species-specific responses outlined elsewhere, and the differences shown in figure 1 between methods, are these really comparable with the UPN spring index data based on clones of lilac and honeysuckle? In figure 4, which suite of species are used?

The reviewers all agreed that we needed to include more on the methods. We have added more regarding Figure 1 and 5 as mentioned above and the exact details for Figure 4 can be found in the supplement. And, finally, as mentioned in our response to Reviewer 2, the aim of Figure 1 was to stress that these three methods are not interchangeable and encourage researchers to carefully choose the best method for their study. We have added more information to the references and have added text to hopefully stress these issues (lines 115-129):

Yet, in contrast to our three metrics of spring onset for one site, most FSI work currently ignores variation across functional groups — instead using one metric (often from SI-x data) of spring onset and assuming it applies to the whole community of plants (Allstadt *et al.*, 2015; Marino *et al.*, 2011; Mehdipour & Zurita-Milla, 2017; Peterson & Abatzoglou, 2014). The risk of a false spring varies across habitats and with species composition since spring onset is not consistent across functional groups (Martin *et al.*, 2010). Therefore, one spring onset date cannot be used as an effective proxy for all species and the three methodologies are not interchangeable.

Rather than use any metric to evaluate spring onset, we encourage researchers to first assess the forest demographics and functional groups relevant to the study question. From there, researchers can choose the most appropriate method to estimate the date of budburst to determine if a false spring could have occurred. However, as we outline below, considering different functional groups is unlikely to be enough for robust predictions in regards to level of damage from a false spring, especially when trying to determine how frosts are shaping the life history of plants. In such cases, it will also be important to integrate species differences within functional groups and to consider the various interspecific tolerance and avoidance strategies that species have evolved against false springs.

References

- Allstadt AJ, Vavrus SJ, Heglund PJ, Pidgeon AM, Wayne E, Radeloff VC (2015) Spring plant phenology and false springs in the conterminous U. S. during the 21st century. *Environmental Research Letters (submitted)*, **10**, 104008. doi:10.1088/1748-9326/10/10/104008. URL <http://dx.doi.org/10.1088/1748-9326/10/10/104008>.
- Augsburger CK (2009) Spring 2007 warmth and frost: Phenology, damage and refoliation in a temperate deciduous forest. *Functional Ecology*, **23**, 1031–1039. doi:10.1111/j.1365-2435.2009.01587.x.
- Augsburger CK (2013) Reconstructing patterns of temperature, phenology, and frost damage over 124 years: Spring damage risk is increasing. *Ecology*, **94**, 41–50. doi:10.1890/12-0200.1.
- Basler D, Körner C (2012) Photoperiod sensitivity of bud burst in 14 temperate forest tree species. *Agricultural and Forest Meteorology*, **165**, 73–81. doi:10.1016/j.agrformet.2012.06.001. URL <http://dx.doi.org/10.1016/j.agrformet.2012.06.001>.
- CaraDonna, Paul J and Bain, Justin A (2016) Frost sensitivity of leaves and flowers of subalpine plants is related to tissue type and phenology. *Journal of Ecology*, **104**, 55–64. doi:10.1111/1365-2745.12482.
- Gu L, Hanson PJ, Post WM, *et al.* (2008) The 2007 Eastern US spring freeze: Increased cold damage in a warming world. *BioScience*, **58**, 253. doi:10.1641/B580311.
- Hufkens K, Friedl MA, Keenan TF, Sonnentag O, Bailey A, O’Keefe J, Richardson AD (2012) Ecological impacts of a widespread frost event following early spring leaf-out. *Global Change Biology*, **18**, 2365–2377. doi:10.1111/j.1365-2486.2012.02712.x.
- Inouye DW (2000) The ecological and evolutionary significance of frost in the context of climate change. *Ecology Letters*, **3**, 457–463. doi:10.1046/j.1461-0248.2000.00165.x.
- Kim Y, Moorcroft PR, Aleinov I, Puma MJ, Kiang NY (2015) Variability of phenology and fluxes of water and carbon with observed and simulated soil moisture in the Ent Terrestrial Biosphere Model (Ent TBM version 1.0.1.0.0). *Geoscientific Model Development*, **8**, 3837–3865. doi:10.5194/gmd-8-3837-2015. URL <http://dx.doi.org/10.5194/gmd-8-3837-2015>.
- Klosterman S, Hufkens K, Richardson AD (2018) Later springs green-up faster: the relation between onset and completion of green-up in deciduous forests of North America. *International Journal of Biometeorology*. doi:10.1007/s00484-018-1564-9. URL <https://doi.org/10.1007/s00484-018-1564-9>.
- Lenz A, Hoch G, Vitasse Y, Körner C (2013) European deciduous trees exhibit similar safety margins against damage by spring freeze events along elevational gradients. *New Phytologist*, **200**, 1166–1175. doi:10.1111/nph.12452.
- Marino GP, Kaiser DP, Gu L, Ricciuto DM (2011) Reconstruction of false spring occurrences over the southeastern United States, 1901–2007: an increasing risk of spring freeze damage? *Environmental Research Letters*, **6**, 24015. doi:10.1088/1748-9326/6/2/024015.
- Martin M, Gavazov K, Körner C, Hattenschwiler S, Rixen C (2010) Reduced early growing season freezing resistance in alpine treeline plants under elevated atmospheric CO₂. *Global Change Biology*, **16**, 1057–1070. doi:10.1111/j.1365-2486.2009.01987.x. URL <http://dx.doi.org/10.1111/j.1365-2486.2009.01987.x>.

- Mehdipoor H, Zurita-Milla EIVR (2017) Continental-scale monitoring and mapping of false spring: A cloud computing solution. URL <http://www.geocomputation.org/2017/papers/48.pdf>.
- Menzel A, Helm R, Zang C (2015) Patterns of late spring frost leaf damage and recovery in a European beech (*Fagus sylvatica* L.) stand in south-eastern Germany based on repeated digital photographs. *Frontiers in Plant Science*, **6**, 110. doi:10.3389/fpls.2015.00110. URL <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC4338663/>.
- Moorcroft PR, Hurtt GC, Pacala SW (2001) A method for scaling vegetation dynamics: The Ecosystem Demography Model (ED). *Ecological Monographs*, **71**, 557–585. URL <http://www.jstor.org/stable/3100036>.
- NOAA (2017) Climate data online search. URL <https://www.ncdc.noaa.gov/cdo-web/search?datasetid=GHCND>.
- Peterson AG, Abatzoglou JT (2014) Observed changes in false springs over the contiguous United States. *Geophysical Research Letters*, **41**, 2156–2162. doi:10.1002/2014GL059266.
- Richardson AD (2018) Tracking seasonal rhythms of plants in diverse ecosystems with digital camera imagery. *New Phytologist*. doi:10.1111/nph.15591. URL <http://dx.doi.org/10.1111/nph.15591>.
- Richardson AD, Hufkens K, Milliman T, Froking S (2018) Intercomparison of phenological transition dates derived from the PhenoCam Dataset V1.0 and MODIS satellite remote sensing. *Scientific Reports*, **8**. doi:10.1038/s41598-018-23804-6. URL <http://dx.doi.org/10.1038/s41598-018-23804-6>.
- Richardson AD, Keenan TF, Migliavacca M, Ryu Y, Sonnentag O, Toomey M (2013) Climate change, phenology, and phenological control of vegetation feedbacks to the climate system. *Agricultural and Forest Meteorology*, **169**, 156 – 173. doi:<https://doi.org/10.1016/j.agrformet.2012.09.012>. URL <http://www.sciencedirect.com/science/article/pii/S0168192312002869>.
- Sakai A, Larcher W (1987) *Frost Survival of Plants*. Springer-Verlag.
- Vitasse Y, Lenz A, Hoch G, Körner C (2014) Earlier leaf-out rather than difference in freezing resistance puts juvenile trees at greater risk of damage than adult trees. *Journal of Ecology*, **102**, 981–988. doi:10.1111/1365-2745.12251.
- Xie Y, Wang X, Silander JA (2015) Deciduous forest responses to temperature, precipitation, and drought imply complex climate change impacts. *Proceedings of the National Academy of Sciences*, **112**, 13585–13590. doi:10.1073/pnas.1509991112. URL <http://dx.doi.org/10.1073/pnas.1509991112>.