

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 adjusting 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 under a new subsection, *Defining false spring: When are plants vulnerable to frost damage?* (lines 70-112), which includes:

At the level of an individual plant, vulnerability to frost damage varies across tissues and seasonally with plant development. Some tissues are often more or less sensitive to low temperatures. Flower and fruit tissues, are often easily damaged by freezing temperatures (Augsburger, 2009; CaraDonna & Bain, 2016; Inouye, 2000; Lenz *et al.*, 2013), while wood and bark tissues can survive lower temperatures through various methods (Strimbeck *et al.*, 2015). Similar to wood and bark, leaf and bud tissues can often survive lower temperatures without damage (Charrier *et al.*, 2011). However, for most tissues, tolerance of low temperatures varies seasonally with the environment through the development of cold hardiness (i.e. freezing tolerance), which allows plants to survive colder winter temperatures through various physiological mechanisms (Sakai & Larcher, 1987; Strimbeck *et al.*, 2015, e.g., deep supercooling, increased solute concentration, and an increase in dehydrins and other proteins). Additionally, plants will suffer greater long-term effects from the loss of photosynthetic tissue through impacts on multiple years of growth, reproduction, and canopy development (Vitasse *et al.*, 2014; Xie *et al.*, 2015). For these reasons, we 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 to Massachusetts and agree this is a crucial point to mention. We added text (lines 157-160) in the *Defining false spring: Measuring false spring in one temperate plant community* section:

While the SI-x data (based on observations of early-active shrub species, especially including the — non-native to Massachusetts — species lilac, *Syringa vulgaris*) may best capture understory dynamics, the PhenoCam and observational FSI data integrate over larger canopy species, which budburst later and thus are at generally lower risk of false springs.

And again in a new section *The future of false spring research* (lines 308-364):

Many studies in particular use gridded spring-onset data (e.g., SI-x). Studies aiming to forecast false spring risk across a species' range using SI-x data may

do well for species similar to lilac (*Syringa vulgaris*), such as other closely related shrub species distributed across or near lilac's native southwestern European range. But we expect predictions would be poor for less similar 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 70-112) as aforementioned with Reviewer 1's comments (see above for some of the relevant text).

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 making substantial changes throughout the paper. We changed the names of many headings and sub-headings and also moved 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 *Integrating phenological cues to predict vegetative risk* a sub-heading under *Improving false spring definitions* and added text to explain that it could also be an avoidance strategy against false springs (lines 226-240):

Understanding what determines the rate of budburst and the length of time between budburst and leafout is essential for predicting the level of damage from a false spring event. The timing between these phenophases (budburst to leafout), which we refer to as the duration of vegetative risk (Figure 3) is a critical area of future research. Currently research shows there is significant variation across species in their durations of vegetative risk, but basic information, such as whether early-budburst species and/or those with fewer morphological traits

to avoid freeze damage have shorter durations of vegetative risk compared to other species, is largely unknown but important for improved forecasting. With spring advancing, species that have shorter durations of vegetative risk would avoid more false springs compared to those that have much longer durations of vegetative risk, especially among species that budburst early. This hypothesis, however, assumes the duration of vegetative risk will be constant with climate change, which seems unlikely as both phenophases are shaped by environmental cues. The duration of vegetative risk is therefore best thought of as a species-trait with potentially high variation determined by environmental conditions. Understanding the various physiological and phenological mechanisms that determine budburst and leafout across species will be important for improved metrics of false spring, especially for species- and/or site-specific studies.

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 address these issues (lines 131-146):

To demonstrate how the FSI definition works—and is often used—we applied it to data from the Harvard Forest Long-term Ecological Research program in Massachusetts. We selected this site as it has been well monitored for spring phenology through multiple methods for several years. While at the physiological level, frost damage is most likely to occur between budburst and leafout, data on the exact timing of these two events are rarely available and surrogate data are often used to capture 'spring onset' (i.e., initial green-up) at the community level. We applied three commonly used methods to calculate spring onset: long-term ground observational data (O'Keefe, 2014), PhenoCam data (Richardson, 2015), and USA National Phenology Network's (USA-NPN) Extended Spring Index (SI-x) "First Leaf - Spring Onset" data (USA-NPN, 2016). These three methods for spring onset values require different levels of effort and are—thus—variably available for other sites. The local ground observational data (O'Keefe, 2014)—available at few sites—requires many hours of personal observation, but comes the closest to estimating budburst and leafout dates. PhenoCam data requires only the hours to install and maintain a camera observing the canopy, then process the camera data to determine canopy color dynamics over seasons and years. Finally, SI-x data can be calculated for most temperate sites, as the index was specifically designed to help provide an available, comparable estimate of spring onset across sites. Once calculated for this particular site we inputted our three estimates of spring onset into the FSI equation (Equation 1) to determine the FSI from 2008 to 2014 (Figure 2).

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 (lines 24-27; 62-63; 175-182).

We have also added a new section *The future of false spring research* to address this more directly in lines 334-339 and lines 350-356 (see also our response to Reviewer 3 regarding this issue):

Related to range studies, studies of plant life history will benefit from more specialized metrics of false spring. Estimates of fitness consequences of false springs at the individual- population- or species-levels must integrate over important species, population and life-stage variation. In such cases, careful field observational and lab experimental data will be key. Through such data, researchers can capture the variations in temperature thresholds, species- and lifestage-specific tolerance and avoidance strategies and climatic effects, and more accurately measure the level of damage. ... While we suggest most studies at the individual to community levels need far more complex metrics of false spring to make major progress, simple metrics of false spring may be appropriate for a suite of studies at ecosystem-level scales. Single-metric approaches, such as the FSI, are better than not including spring frost risk in relevant studies. Thus, these metrics could help improve many ecosystem models, including land surface models (Foley *et al.*, 1998; Moorcroft *et al.*, 2001; Prentice *et al.*, 1992). In such models, the SI-x combined with FSI would provide researchers with predicted shifts in frequency of false springs under emission scenarios.

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 288-296) as well as reference the supplement in the caption:

To highlight this, we analyzed five archetypal regions across North America and Europe. Through the use of both phenology (Soudani *et al.*, 2012; Schaber & Badeck, 2005; USA-NPN, 2016; White *et al.*, 2009) and climate data (from the NOAA Climate Data Online tool NOAA, 2017) we determined the number of false springs (i.e., temperatures at -2.2°C or below) for each region. Here, we used the FSI equation and tallied the number of years when FSI was positive. 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). Here FSI was a valuable resource to elucidate the regional differences in false spring risk, but for useful projections these estimates should be followed up with more refined data (see *The future of false spring research* below).

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 207-211):

Species that budburst late are expected to have high requirements of chilling, forcing and/or photoperiod. For example, the combination of a high chilling and a spring forcing requirement (that is, a species which requires long periods of cool temperatures to satisfy a chilling requirement before responding to any forcing conditions) will avoid budbursting during periods of warm temperatures too early due to insufficient chilling (Basler & Körner, 2012).

Line 36: this research -> these researches.

We agree there are issues with this sentence. We have now improved it by removing the redundancies (see lines 55-57).

To produce accurate predictions, researchers need improved methods that can 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 text throughout (see lines 24-27; 62-63), more to the end of section *Defining false spring: Measuring false spring in one temperate plant community* (lines 175-182) and a new section *The future of false spring research* (see lines 307-367), which includes:

Ideally researchers should first assess the forest demographics and functional groups relevant to their study question, then select the most appropriate method to estimate the date of budburst to determine if a false spring could have occurred. This, however, still ignores variation in the date of leafout (when cold

tolerance increases slightly). Further, considering different functional groups is unlikely to be enough for robust predictions in regards to level of damage from a false spring, especially for ecological questions that operate at finer spatial and temporal scales. For many research questions—as we outline below—it will be important to develop false spring metrics that integrate species differences within functional groups, by considering the tolerance and avoidance strategies that species have evolved to mitigate false springs effects.

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: When are plants vulnerable to frost damage?* (see lines 70-112).

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 greatly 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 (see lines 24-27; 62-63; 175-182; 297-306). We have also added an entirely new section *The future of false spring research* (lines 340-343 and 350-356) to address these concerns more directly (see also our response regarding this to Reviewer 2):

Though time-consuming, we suggest research to discover species x life-stage specific freezing tolerances and related cues determining the duration of vegetative risk to make major advances in fundamental and applied science. Such studies can help determine at which life stages false springs have important fitness consequences, and whether tissue damage from frost for some species x life stages actually scales up to minimal fitness effects. ... While we suggest most studies at the individual to community levels need far more complex metrics of false spring to make major progress, simple metrics of false spring may be appropriate for a suite of studies at ecosystem-level scales. Single-metric approaches, such as the FSI, are better than not including spring frost risk in relevant studies. Thus, these metrics could help improve many ecosystem models, including land surface models (Foley *et al.*, 1998; Moorcroft *et al.*, 2001; Prentice *et al.*, 1992; Thornton *et al.*, 2005). In such models, the SI- x combined with FSI would provide researchers with predicted shifts in frequency of false springs under emission scenarios.

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 under a new heading *The future of false spring research*.

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 originally included these details in the supplement but we failed to make this clear in the main text or figure captions. 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, which is highlighted in the figure caption. 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 changed the text about the three methods used to determine FSI and have added more information to the references to better address these issues (lines 166-182):

Differing FSI estimated from our three metrics of spring onset for the same site and years highlight variation across functional groups, which FSI work currently ignores — instead using one metric of spring onset (often from SI-x data, which is widely available) 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). As the risk of a false spring varies across habitats and functional groups (Martin *et al.*, 2010) one spring onset date cannot be used as an effective proxy for all species and researchers should more clearly align their study questions and methods. FSI using such estimates as the SI-x may discern large-scale basic trends across space or years, but require validation with ground observations to be applied to any particular location or functional group of species.

Ideally researchers should first assess the forest demographics and functional groups relevant to their study question, then select the most appropriate method to estimate the date of budburst to determine if a false spring could have occurred. This, however, still ignores variation in the date of leafout (when cold tolerance increases slightly). Further, considering different functional groups is unlikely to be enough for robust predictions in regards to level of damage from a false spring, especially for ecological questions that operate at finer spatial

and temporal scales. For many research questions—as we outline below—it will be important to develop false spring metrics that integrate species differences within functional groups, by considering the tolerance and avoidance strategies that species have evolved to mitigate false spring effects.

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.
- 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.
- 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.
- CaraDonna PJ, Bain JA (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.
- Charrier G, Bonhomme M, Lacointe A, Améglio T (2011) Are budburst dates, dormancy and cold acclimation in walnut trees (*Juglans regia* L.) under mainly genotypic or environmental control? *International Journal of Biometeorology*, **55**, 763–774. doi:10.1007/s00484-011-0470-1.
- Foley JA, Levis S, Prentice IC, Pollard D, Thompson SL (1998) Coupling dynamic models of climate and vegetation. *Global Change Biology*, **4**, 561–579. doi:10.1046/j.1365-2486.1998.t01-1-00168.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.
- 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 CO_2 . *Global Change Biology*, **16**, 1057–1070. doi: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>.
- 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>.
- O’Keefe J (2014) Phenology of woody species at Harvard Forest since 1990. URL <http://harvardforest.fas.harvard.edu:8080/exist/apps/datasets/showData.html?id=hf003>.

- 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.
- Prentice IC, Cramer W, Harrison SP, Leemans R, Monserud RA, Solomon AM (1992) Special paper: a global biome model based on plant physiology and dominance, soil properties and climate. *Journal of biogeography*, pp. 117–134.
- Richardson AD (2015) PhenoCam images and canopy phenology at Harvard Forest since 2008. URL <http://harvardforest.fas.harvard.edu:8080/exist/apps/datasets/showData.html?id=hf158>.
- Sakai A, Larcher W (1987) *Frost Survival of Plants*. Springer-Verlag.
- Schaber J, Badeck FW (2005) Plant phenology in Germany over the 20th century. *Regional Environmental Change*, **5**, 37–46. doi:10.1007/s10113-004-0094-7.
- Soudani K, Hmimina G, Delpierre N, *et al.* (2012) Ground-based Network of NDVI measurements for tracking temporal dynamics of canopy structure and vegetation phenology in different biomes. *Remote Sensing of Environment*, **123**, 234–245. doi:10.1016/j.rse.2012.03.012.
- Strimbeck GR, Schaberg PG, Fossdal CG, Schröder WP, Kjellsen TD (2015) Extreme low temperature tolerance in woody plants. *Frontiers in Plant Science*, **6**. doi:10.3389/fpls.2015.00884.
- Thornton PE, Running SW, Hunt ER (2005) Biome-bgc: Terrestrial ecosystem process model, version 4.1.1. doi:10.3334/ornl daac/805.
- USA-NPN (2016) USA National Phenology Network Extended Spring Indices. URL <http://dx.doi.org/10.5066/F7XD0ZRK>.
- 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.
- White MA, De Beurs KM, Didan K, *et al.* (2009) Intercomparison, interpretation, and assessment of spring phenology in North America estimated from remote sensing for 1982–2006. *Global Change Biology*, **15**, 2335–2359. doi:10.1111/j.1365-2486.2009.01910.x.
- 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.