Intro: ① Late spring freezes / introduce false spring concept
  (a) current opening sentence
       (Indiv?)
  (b) Spp. that leaf out before the last frost
      ~~often~~ are at risk of damaging ~~used~~ tissues,
      leaf loss and slowed or stalled canopy
      development (Gu et al 2008, Hufkens et al. 2012).
  (c) Some sort of wrap up sentence
      suggesting plants should avoid
      this or such OR define
      (simplistically) false spring.

↑ in Richardson's lab, right? Have you so. Chatted? If not, you should!

② False spring & climate change (CC) ↓plant
  (a) CC is expected to increase damage from false spring
      events around the world. (or something simple) ᵥsimilarly
  (b) earlier springs (sentence you have)
  (c) ~~Clim~~ Multiple studies have documented false spring events
      in recent yrs (CITES) + some have linked this to
      CC. (cites).
  (d) Some sort of concluding / transition sentence

③ ┌Current research (overview of) + its simplifications
   → Organize some of what you have in intro into a very
     quick overview of FSI + major research
   → Issues w/ these simplifications (1-2 sentences max)
   → In contrast to these simplifications (we argue that a
     wealth of factors greatly impacts plants' frost spring
     risk such that simple indices will (do little to
     advance the field // lead to inaccurate predictions//
     or what?)

④ In this paper we aim ...
     highlight?
   - to ‹unpack› the complexity of factors driving a plant's
     false spring risk
   - We outline in particular how X,Y,Z major factors
     alter / unhinge simple metrics of false spring ~~risks~~
  - and show how an integrated view of false spring that
    incorporates these factors would rapidly advance progress
    in this field.

Section 2 Defining false Spring (DFS)

- See DFS notes throughout

- I think "Determining Spring Onset" can go here but you need to develop idea 1st (e.g. canopy vs. understory, sp. vs. mix etc.) ... then results only + digest what they mean for reader

Section 3 Understanding (Defining?) Vegetative risk

I think you could also fit...
(see [a] below)

→ Phenophases
→ Spp. differences (maybe just set this up here + ...)
→ regional diffs?

Section 4 Spp. differences + vegetative risk

not sure of order of if your want X more sections.

You have so much great stuff here but need to develop WHY + HOW spp. would differ. You touch on this once in paper but need to develop more. This section should start w/ building a case for why spp. would differ then go into data.

Section 5 Regional differences in veg. risk?

→ Again need to build up for readers why regions might differ + How (basically get your readers to come up w/ your hypothesis before you spell it out) then
(1) analyze lit for regional diffs
(2) latitude analysis

and maybe a few other pts you make here fit

[a] continued: some of your points in Box 1 here — basically set up everything that could matter then make a case for what matters most + hit on those in next sections.

Section 6 Conclusions.

1 Introduction

Plants that grow in temperate environments are at risk of being exposed to late spring freezes, which can be detrimental to plant growth. According to Gu et al. 2008, there are two phases involved in late spring freezing: rapid vegetative growth prior to the freeze and the post freeze setback. This combined process is known as a false spring. Freeze and thaw fluctuations can cause xylem embolism and decreased xylem conductivity which can result in crown dieback (Gu *et al.*, 2008). More frequently, plants that have been exposed to a false spring will experience leaf loss and slower canopy development (Hufkens *et al.*, 2012). Spring frosts during the vegetative growth phenophases impose the greatest freezing threat to deciduous tree species (Sakai & Larcher, 1987). With anthropogenic climate change, the severity of damage incurred from a false spring phenomena is predicted to be heightened due to earlier spring onset and greater fluctuations in temperatures. Temperate forest species around the world are initiating leaf out about 4.6 days earlier per degree Celsius (Polgar *et al.*, 2014; Wolkovich *et al.*, 2012). It is anticipated that there will be a decrease in false spring frequency overall but the severity of temperature variation is likely to increase, therefore amplifying the expected intensity of false spring events (Allstadt *et al.*, 2015). In 2012, a false spring event in Michigan resulted in half a billion dollars worth of fruit tree damage (Ault *et al.*, 2013; Knudson, 2012). Due to these reasons, it is crucial for researchers to properly evaluate the effects of false spring events on not only agricultural crops but in temperate forests as well in order to make more accurate predictions on future trends.

Different species respond differently to late spring freezing events. The level of damage sustained by plants from a false spring also varies across phenophases. Generally, reproductive phases are more sensitive to false spring events than vegetative phases and developing leaves are more susceptible to damage than opening buds or expanding shoots (Lenz *et al.*, 2013; Augspurger, 2009). However, trees that suffer severe vegetative growth damage from a false spring event will suffer greater long-term effects from the loss of photosynthetic tissue than trees that lose one year of reproductive growth. False spring events put seedlings and saplings at greater risk to damage than adult trees (Vitasse *et al.*, 2014). Younger trees are more likely to incur lastly damage to the leaf buds and vegetative growth, whereas adult trees are at risk of xylem embolism. In order for xylem embolism to occur, extreme cavitation must first occur. Extensive cavitation in the xylem would require more intensive freezing events than it would take to damage seedling and sapling leaf buds. Especially strong freezing events (i.e. >-8.6°C), could result in meristemic tissue, wood parenchyma and phloem damage (Lenz *et al.*, 2013; Augspurger, 2011; Sakai & Larcher, 1987). In a study performed by Augspurger 2009, it was noted that some individuals of the same life stage were at different phenophases during a freezing event and subsequently each individual suffered varying degrees of damage. Thus, indicating that phenophase is a greater indicator for level of risk than life stage. Warm temperatures earlier in the year (i.e. in February) do

4

not seem to affect species, most likely because it is too soon for budburst to initiate and sufficient chilling has not yet occurred. Frost damage usually occurs when there is a warmer than average March, a freezing April, and enough growing days between the high temperatures and the last freeze date (Augspurger, 2013). In a study performed by Peterson and Abatzoglou 2014, it had been determined that 7 days between budburst and last freeze date is a significant parameter. During this time, it was determined that leaf buds will have enough growing degree days to begin budburst but the leaves won't have fully expanded yet. There is much debate over the definition of freezing temperatures, which has resulted in two types of freezes: a "hard" freeze at -2.2°C and a "soft" freeze at -1.7°C (Augspurger, 2013; Kodra *et al.*, 2011; Vavrus *et al.*, 2006).

At this time false spring studies fail to incorporate all potential factors that could affect the level of frost damage risk. A False Spring Index (FSI) signifies the likelihood of a damage to occur from a late spring freeze. Currently, FSI evaluates day of budburst, number of growing degree days, and day of last spring freeze through a simple equation as seen below (Marino *et al.*, 2011).

$$FSI = JulianDate(LastSpringFreeze) - JulianDate(Budburst)$$

If FSI is a positive number and greater than 7, then crown dieback is more likely to occur. In this study, we aim to integrate a more thorough model for predicting false spring risk, which would ideally incorporate life stage of the individual (Caffarra & Donnelly, 2011), location within a forest or canopy (Augspurger, 2013), winter chilling hours (Flynn & Wolkovich 2017?), proximity to water (Gu *et al.*, 2008), level of precipitation prior to the freezing event (Anderegg *et al.*, 2013), freeze duration/intensity, and range limits of the species. Another highly crucial factor to consider is the rate of budburst and the length of time between budburst to full leaf out, which we will refer to as the duration of vegetative risk. Temperate trees are most susceptible to damage during leaf out and expansion (Vitasse *et al.*, 2014). We will use the BBCH Scale Phase 09 to define budburst and Phase 15 to define leaf out (Meier, 2001). In a study by Lenz et al. 2013, it was determined that elevation is not a key indicator for determining the level of risk from false springs. Likewise, wood density is also not a significant parameter (Augspurger, 2009). For these reasons, elevation and wood density will be excluded from the suggested model.

2 Climatic Implications: How climate change is affecting spring freezing events

Temperate deciduous tree species must have plastic phenological responses in the spring in order to optimize photosynthesis and minimize frost or drought risk (Polgar & Primack, 2011). In a study performed by Vitasse et al. 2013, environmental effects (i.e. temperature) had a greater effect on seedling flushing than genetic

effects. However, different species respond differently to anthropogenic climate change. Most species are expected to begin leaf out earlier in the season with warming spring temperatures but some species may have the opposite response (Xin, 2016; Cleland *et al.*, 2006; Yu *et al.*, 2010).

Studies indicate that species growing at more northern latitudes tend to respond greater to photoperiod than species growing further south (Caffarra & Donnelly, 2011). Similarly, late successional species exhibit greater photoperiod sensitivities than pioneer species (Basler & Körner, 2012). Based on this information, we analyzed two latitudinal gradients by downloading Daily Summary climate datasets from the NOAA Climate Data Online tool. We assessed 8-10 different degree latitude lines for each transect in order to measure frequency of false spring events (Menne *et al.*, 2012; Menne *et al.*). False springs were tallied by first calculating the number of Growing Degree Days (GDD) using the equation formulated by Michigan State University as seen below (Nugent, J., 2005).

$$DegreeDays = AverageDailyTemp(°C) - 10°C$$

If there were 40 GDDs before a hard freeze occurred in the spring (-2.2°C), then it was determined that a false spring could have occurred in that year. Since we did not incorporate actual budburst or spring onset information, it is unclear whether these events were actually damaging. In order to simply address the climate question, we used these parameters to have a better understanding of the potential climate effects of latitude. Each location includes 30 years of climate data and each transect fell within 3 degrees longitude. Locations that were over 1,000m above sea level were excluded. Table 1 shows the results from the European transect investigated. False spring occurrence ranged from 0 to 8 and increased as latitude decreased.

Table 1: Number of False Springs along a Latitudinal Gradient in Western Europe

Station	Elevation	Latitude	Longitude	False Springs
Kempten, Germany	705m	47.724	10.336	8
Augsburg, Germany	461m	48.426	10.943	8
Bamberg, Germany	210m	49.875	10.921	7
Jena, Germany	155m	50.927	11.584	4
Hannover, Germany	55.0m	52.466	9.679	6
Bremen, Germany	4.00m	53.046	8.799	5
Hamburg, Germany	11.0m	53.635	9.99	4
Schleswig, Germany	43.0m	54.529	9.549	0
Flyvestation, Denmark	3.00m	57.093	9.849	0
Oslo, Norway	94.0m	59.943	10.721	0

Table 2 shows the results from the American transect. False spring occurrence ranged from 0 to 13 and also

6

exhibited an inverse relationship with latitude.

Table 2: Number of False Springs along a Latitudinal Gradient in North America

Station	Elevation	Latitude	Longitude	False Springs
Anthony, Kansas	415m	37.155	-98.028	13
Hastings, Nebraska	587m	40.583	-98.350	7
West Point, Nebraska	399m	41.845	-96.714	5
Yankton, South Dakota	360m	42.883	-97.350	5
Brookings, South Dakota	497m	44.325	-96.769	0
Aberdeen, South Dakota	395m	45.443	-98.413	2
Grand Forks, North Dakota	253m	47.933	-97.083	0
Pembina, North Dakota	241m	48.971	-97.242	1

Table 4 shows the results from the linear regression models performed for both transects. Latitude (1) represents the European transect and Latitude (2) is the American transect.

As seen in the above tables, as latitude increased the frequency of false spring events decreased. These results may indicate why species with a more northern range may have greater photoperiod sensitivities because the level of risk associated with a frost may in fact decrease. These findings demonstrate that further research needs to be pursued but they could ultimately indicate that certain latitudes should be prioritized for future false spring studies.

3   Determining Spring Onset

Before a suitable model for determining false spring risk can be established, an appropriate determination of spring onset is crucial. There are many methods that can be used to determine first day of spring and there are also many definitions. In order to test the best technique for calculating spring onset (or budburst), we gathered data using three different methodologies. The first method for collecting budburst was from observational data recorded for 33 tree species by Dr. John O'Keefe at Harvard Forest from 1990 to 2014 (O'Keefe, 2014). Dr. O'Keefe defines budburst as 50% green tip emergence. We subsetted this dataset to include only the tree species that were most consistently observed, which ended up being eight species. The second dataset was from PhenoCam data, which are field cameras placed in the Harvard Forest canopy that take real-time images of plant growth and are programmed to record initial green up. The final set was collected through the USA National Phenology Network (USA-NPN), using their Data Visualization tool to gather Extended Spring Index values (SI-x) by accessing the "Spring Indices, Historic Annual" gridded layer

Supercool

Table 3: The results from a linear regression model analyzing the relationship between latitude and frequency of false springs

	Dependent variable:	
	Latitude	
	(1)	(2)
False.Springs	−1.064***	−0.801***
	(0.180)	(0.148)
Constant	57.235***	46.946***
	(0.936)	(0.865)
Observations	10	8
R^2	0.813	0.830
Adjusted R^2	0.790	0.801
Residual Std. Error	1.743 (df = 8)	1.732 (df = 6)
F Statistic	34.885*** (df = 1; 8)	29.251*** (df = 1; 6)

Note: *p<0.1; **p<0.05; ***p<0.01

and looking specifically at "First Leaf - Spring Onset" (USA-NPN, 2016a). The SI-x value uses the time of leaf out using historical dates of budburst from honeysuckle and lilac clones around the U.S. and combines that information with daily recordings from local weather stations (USA-NPN, 2016b; Ault *et al.*, 2015b,a; Schwartz *et al.*, 2013; Schwartz, 1997). Through assessing past years' weather and budburst, scientists are able to determine general weather trends that subsequently lead to leaf out. Based on these trends, SI-x values can be calculated from daily weather data (USA-NPN, 2016b).

$ T_{-\xi m \ln \xi} $

The date of last spring freeze was gathered from the Fisher Meteorological Station which was downloaded from the Harvard Forest web page (data available online[1]). The Tmin values were used and the last spring freeze was determined from the latest spring date that the temperature reached -1.7°C or below.

PhenoCam data is not available for Harvard Forest until 2008 and observation data is only recorded through 2014, so this evaluation assesses FSI values from 2008 through 2014. The FSI values were calculated for each methodology using the formula based on the study performed by Marino et al. (2011). Table 2 shows that the Observed and PhenoCam FSI values are all negative from 2008 through 2014, except for 2012 when the observational data indicates a false spring event. The FSI values from the USA-NPN are typically much higher in comparison to the other two methods.

A graphical representation of the FSI values compared across the three methodologies can be seen in Figure 1. In 2008 and 2012, FSI is higher than the significant parameter given of 7 for the NPN data, indicating a possibly damaging false spring event. However, the PhenoCam data did not indicate a false spring in either of the two years and the FSI value for 2012 found through the observed data was not significant.

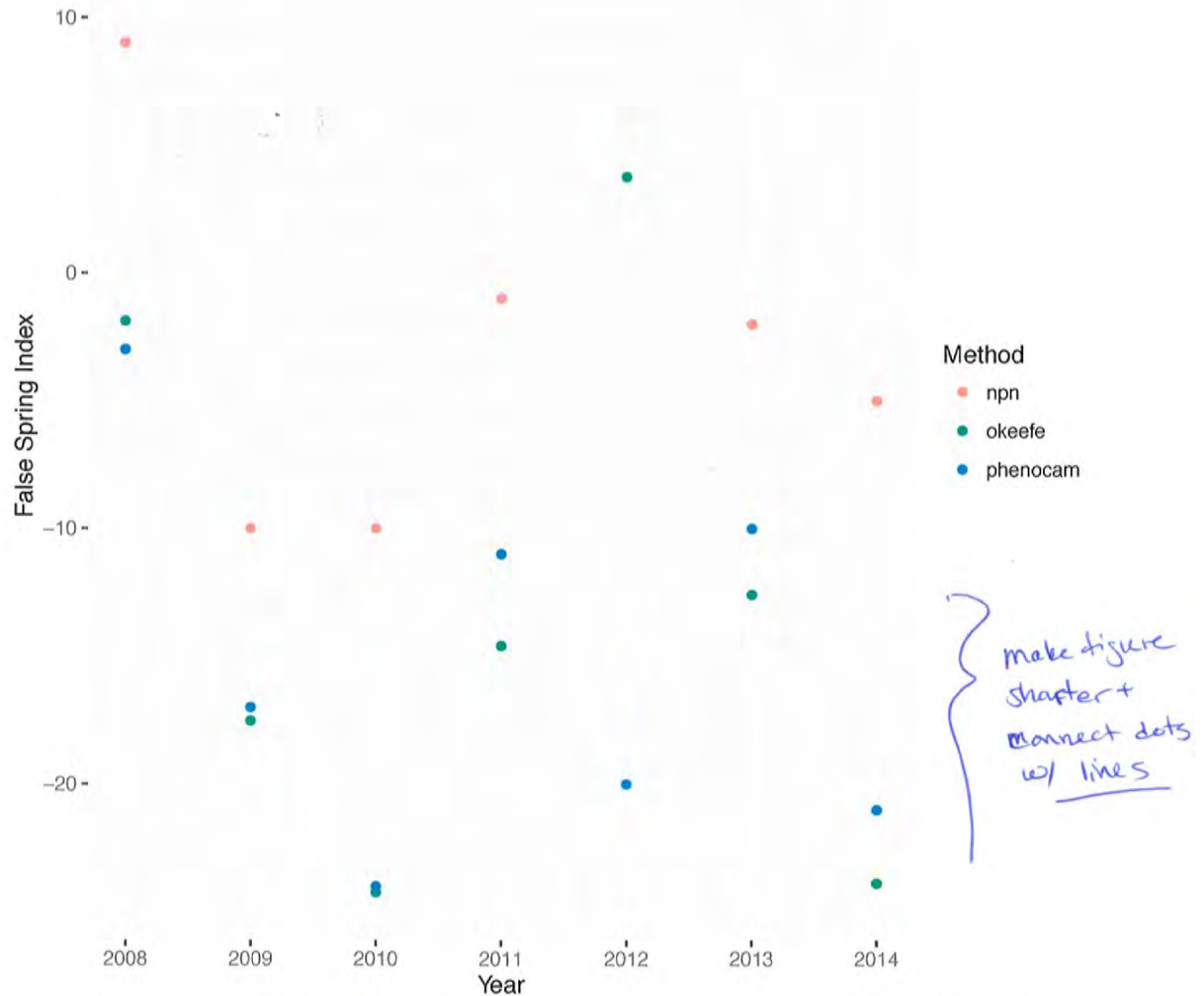[1]http://harvardforest.fas.harvard.edu/meteorological-hydrological-stations

Figure 1: A scatterplot indicating FSI values from 2008 to 2014 for each methdology used in this study. PhenoCam FSI values are red, Observed FSI values are blue, and USA-NPN FSI values are green.

A Pearson Correlation was used to determine the strength of association between the three methods used in the study. As indicated in Table 1, the FSI values from the observed data and the SI-x NPN data are strongly correlated (r=0.9395), whereas the FSI values calculated using the PhenoCam data is not as strongly correlated to either the observed FSI values (r=0.4680) or the NPN FSI values (r=0.4242). Although, according to the Pearson correlation, all methods were considered to be significantly related.

Table 4: Pearson Correlation Coefficients indicating the strength of association between the FSI values calculated across all three methodologies.

	year	npn	okeefe	phenocam
year	1.00	-0.03	-0.20	-0.37
npn	-0.03	1.00	0.94	0.42
okeefe	-0.20	0.94	1.00	0.47
phenocam	-0.37	0.42	0.47	1.00

Our projections indicate that observational FSI values are highly comparable to the USA-NPN FSI values, rendering both justifiable methods for determining potential risk involved in late spring freezes. Even though budburst is defined differently between Dr. O'Keefe, USA-NPN, and the PhenoCam, the dates of budburst are similar. The spring onset dates gathered from the PhenoCam dataset are different from the other two methods, which is likely due to fact that the PhenoCam data is assessing budburst for the forest canopy. Through the use of USA-NPN data, researchers could gather dates of budburst across multiple locations at once in order to determine False Spring risk, making it a more effective method than observational data. Although, all three methods are viable. Various studies have shown that understory species will initiate budburst earlier in the season in order to exploit open canopies and early growth, whereas late successional species may start later in the season to avoid frost or drought risk (Xin, 2016; Richardson & O'Keefe, 2009). Therefore, the methodology used for determining spring onset should largely be dependent on the functional group of interest: researchers should use the USA-NPN dataset for understory species, PhenoCam data for late successional species, and observational data for a wide array of plant functional types.

4 Determining the Duration of Vegetative Risk

In a study performed by Augspurger 2009, the duration of vegetative risk increased if a freezing event occurred during the phenophases between budburst and leaf expansion. In the same study, refoliation was not consistent among damaged versus undamaged individuals. Augspurger believes if the duration of vegetative risk is longer, than the buds and leaves will be heartier against frost. The species that had short durations of vegetative risk sustained higher levels of damage. In a similar study done by Lenz et al. 2016, it was found that after budburst began, frost tolerance steadily decreased, with leaf out being the most susceptible to frost damage until the leaf was fully unfolded. It is therefore crucial that more studies investigate the relationship between false spring events and duration of vegetative risk.

4.1 Tree Spotters Data

to methods

We analyzed a dataset from a USA-NPN citizen science program, the Arnold Arboretum Tree Spotters , to discern the relationship between duration of vegetative risk and initial budburst date. Figure 2 shows the duration of vegetative risk for 11 different species observed at the Arnold Arboretum in 2016. *Quercus alba* and *Betula nigra* had the longest durations of vegetative risk, most likely due to the fact that one is late successional species and the other is northern species respectively. Overall, there is no significant relationship between duration of vegetative risk and day of budburst. This could be due to the fact that there are various functional groups involved in this study, that it is just over the course of one year, that it is in an arboretum, and that it is a citizen science project. Further investigations should be made to gain a better understanding of observational studies for this interaction.
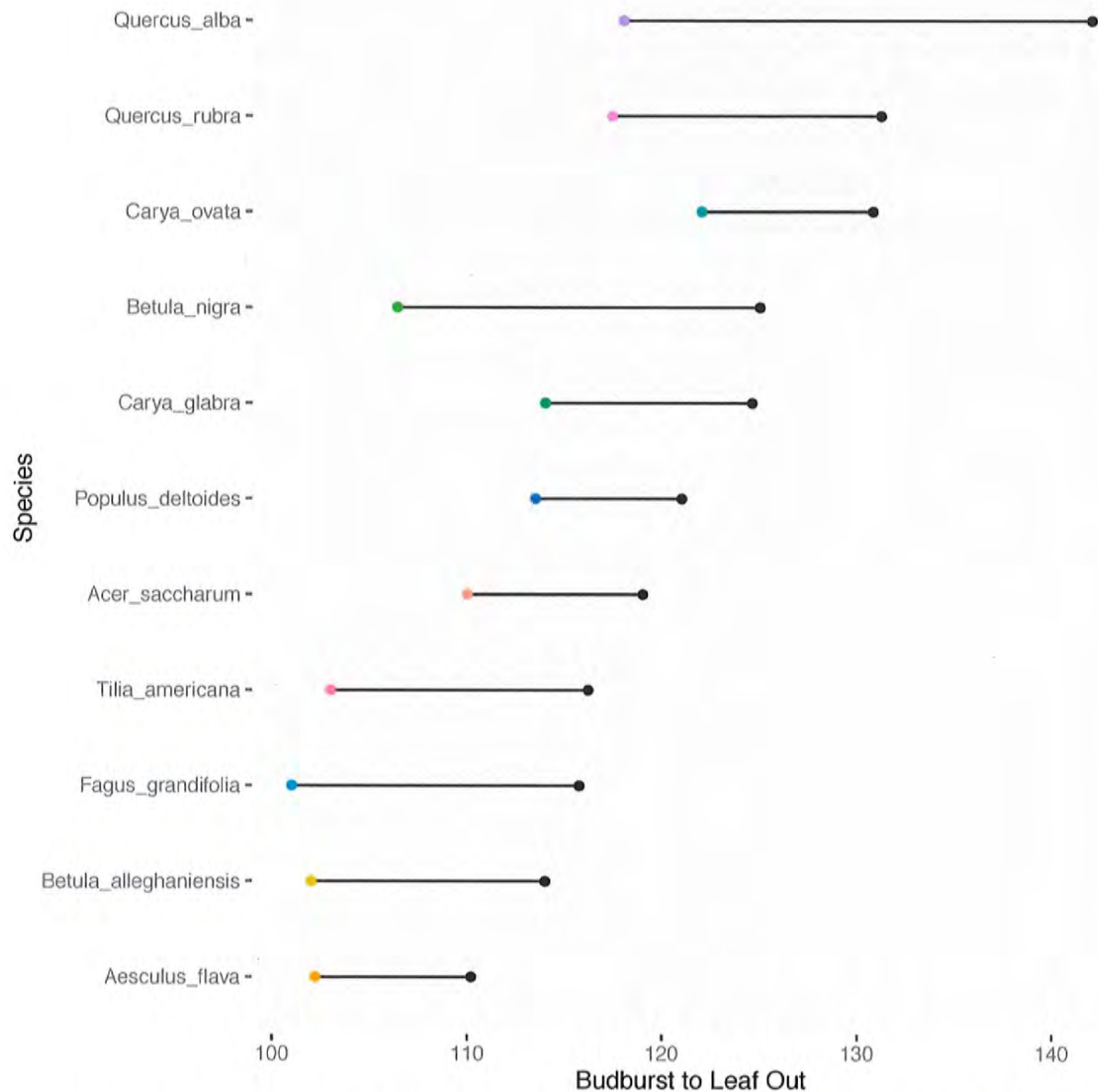
Figure 2: A timeline plot indicating the duration of vegetative risk for each species studied at the Arnold Arboretum in 2016.

(Have you also looked at this for Harvard Forest O'Keefe data?) IGNORE!

4.2 Experimental Chilling Data

It is possible, that with anthropogenic climate change progressing, leaf out timing may be delayed. As winter seasons begin to warm and chilling requirements are not met, more warming in the spring must first occur for budburst to begin (Polgar *et al.*, 2014; Fu *et al.*, 2012; Morin *et al.*, 2009; McCreary *et al.*, 1990). In

I would cite a Chuine paper here. Polgar is just building on Chuine ... (but good to cite Polgar once or twice).

13

a chilling experiment performed by Flynn & Wolkovich (2017), there were various experimental chilling, photoperiod, and forcing treatments. In Figure 2, five species were assessed across five different treatments. C is a forcing temperature of 15°C during the day and 5°C at night, W is a forcing temperature of 20°C during the day and 10°C at night, S is a short day with 8 hours of daylight, L is a long day with 12 hours of daylight, 0 is no additional winter chilling, 1 is 33 days of additional winter chilling at 4°C, and 2 is 33 days of additional winter chilling at 1.5°C. QUERUB is *Quercus rubra*, ACERUB is *Acer rubrum*, POPGRA is *Populus grandidentata*, ILEMUC is *Ilex mucronata*, BETPAP is *Betula papyrifera*.
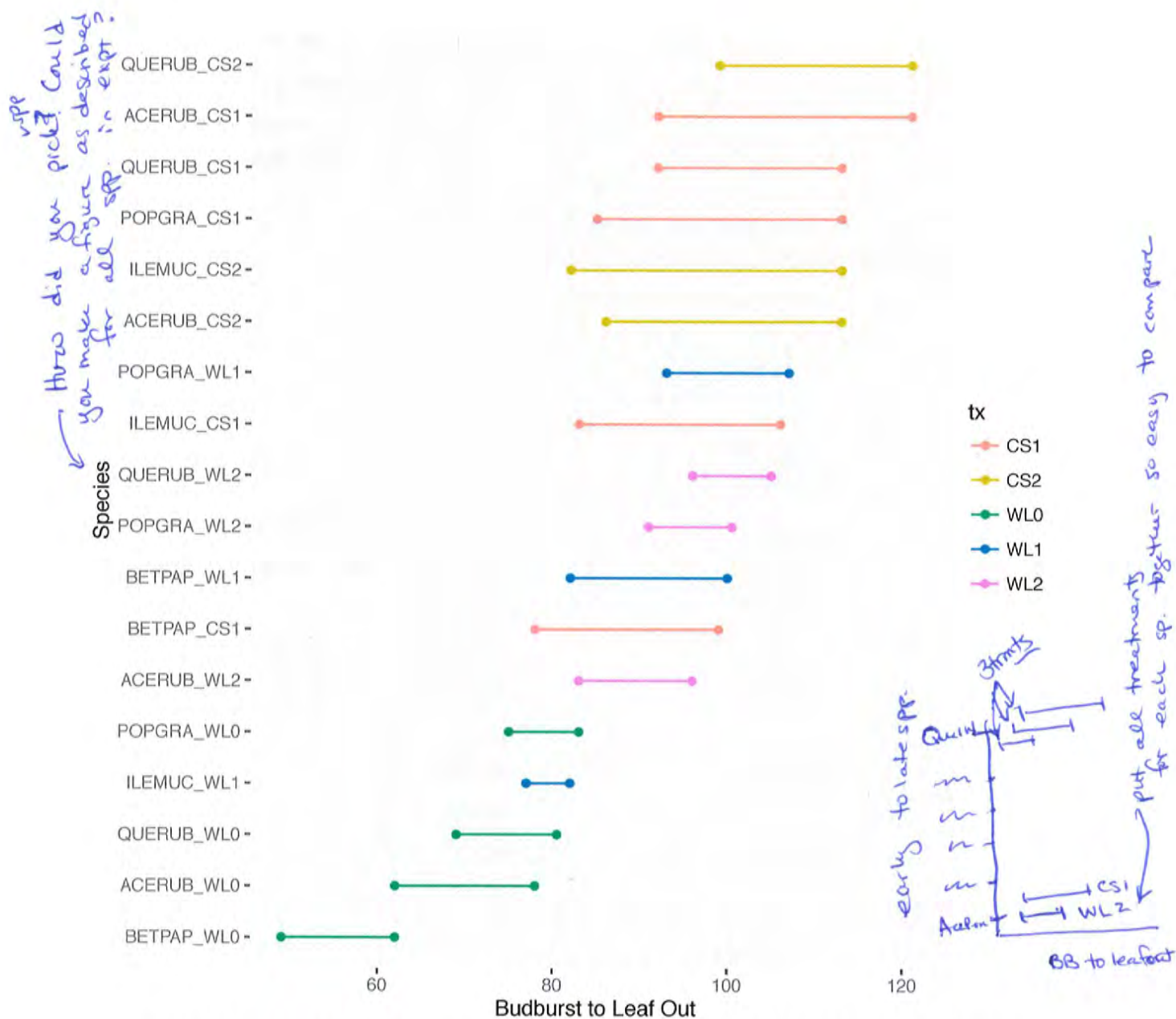
Figure 3: A timeline plot indicating the duration of vegetative risk for each species from experimental chilling study. *(organized by... ?)*

As seen in Table 5, the warmer forcing temperatures (20°C during the day and 5° at night) had the greatest affect on duration of vegetative risk. High forcing temperatures greatly reduced the length of time it took between budburst and leaf out for all species.

Cool! We'll have to discuss how much or little of this to put in so that it doesn't become a data paper.

15

Yes! Nice! I was thinking of this — probably too hard to put in a 'perspective paper' but let's do this analysis.

Table 5: The results from a linear regression model analyzing the relationship between duration of vegetation risk and intial budburst across five treatments

	Dependent variable:
	Risk
Budburst	−0.094
	(0.144)
txCS2	2.550
	(3.179)
txWL0	−14.376***
	(4.315)
txWL1	−12.256***
	(3.163)
txWL2	−13.522***
	(3.202)
Constant	32.522**
	(12.521)
Observations	18
R^2	0.790
Adjusted R^2	0.703
Residual Std. Error	4.313 (df = 12)
F Statistic	9.031*** (df = 5; 12)

Note: *p<0.1; **p<0.05; ***p<0.01

OR:

BB−LO duration ~
~~BB~~ temp × photo × sp

[lots of options? To discuss...]

4.3 Harvard Forest Observational Data

[handwritten margin note: (Ah? Sorry — I really thought you had done this — ignore my earlier comment)]

The final dataset for measuring duration of vegetative risk against initial day of budburst is from John O'Keefe's observational data that was also used in the *Determining Spring Onset* section. For this portion of the study, we assessed two years of data: one year that had an unusually early spring onset (2010) and an unusually late spring (2014). QUAL is *Quercus alba*, FRAM is *Fraxinus americana*, BEAL is *Betula alleghaniensis*, ACRU is *Acer rubrum*, FAGR is *Fagus grandifolia*, ACPE is *Acer pensylvanicum*, QURU is *Quercus rubra*, and HAVI is *Hamamelis virginiana*. As is evident in Figure 4, the duration of vegetative risk is slightly longer in 2010 which was when spring onset was unusually early. Given the disparaging results *[handwritten: disparate?]* across the three studies assessing duration of vegetative risk, more research should be done in order to better understand this relationship.

[handwritten note: Can you say more? What pattern do emerge? What is not true / possibly true? For example — tree? Spp. are different... possibly true — cues matter... Just brainstorm a lot of possibilities + we can discuss which are impt. points to make]
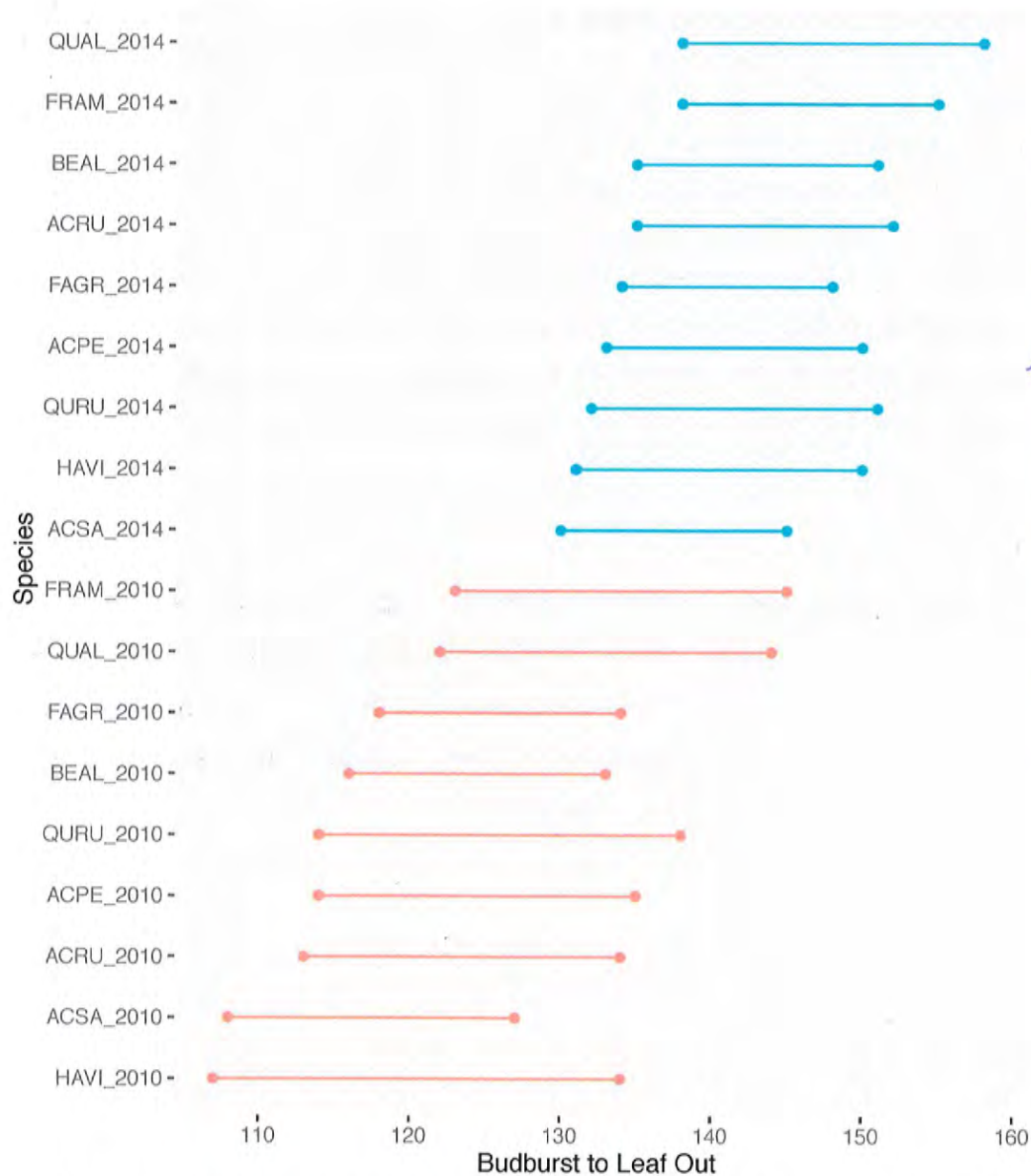
17

Figure 4: A timeline plot indicating the duration of vegetative risk for each species from collected from Harvard Forest.

5 Conclusion

With anthropogenic climate change, false spring risk is higher and the level of damage expected is also greater. Furthermore, habitat fragmentation is increasing. Understanding the impact of false springs on

forest communities, especially along forest edges, would be invaluable. The climatic implications suggest spring forcing temperatures will increase, thus resulting in potentially earlier date of budburst and increased risk for frost or drought damage. However, as is evident from the latitudinal gradients, risk may be greater at certain latitudes. By simply using climate data for this analysis, it shows that the level of risk is simply a regional effect rather than species or functional group specific. The interaction of species and functional group with latitude should be investigated. Calculating spring onset as accurately as possible is essential in order to determine the intensity of a false spring event. Therefore, integrating functional group within a community should be evaluated prior to choosing a method for determining spring onset. Finally, the duration of vegetative risk must be assessed further. The three study types suggest various results with the clearest indication being that greater forcing temperatures in the spring will result in shorter and earlier durations of vegetative risk. Future studies should look at the relationship between phenological plasticity and duration of vegetative risk as well as level of damage in relation to duration of vegetative risk. Box 1 is a list of all the key indicators necessary at this time to properly evalate false spring risk and damage. These indicators should first be tested for significance and then integrated into a model for agricultural and ecological projections.