Comparing spring phenology in an urban arboretum versus rural forested site and implications for forecasting

Authors:

C. J. Chamberlain [1,2] & E. M. Wolkovich [1,2,3]

Author affiliations:

[1] Arnold Arboretum of Harvard University, 1300 Centre Street, Boston, Massachusetts, USA;

[2] Organismic & Evolutionary Biology, Harvard University, 26 Oxford Street, Cambridge, Massachusetts, USA;

[3] Forest & Conservation Sciences, Faculty of Forestry, University of British Columbia, 2424 Main Mall, Vancouver, BC V6T 1Z4

 *Corresponding author: 248.953.0189; cchamberlain@g.harvard.edu

Abstract

Understanding and predicting plant phenology in temperate forests is critical for forecasting important processes such as carbon storage, especially with climate change and urbanization shifting many phases including budburst and leafout. One major forecasting method for phenology is the growing degree day (GDD) model, allows researchers to track heat accumulation. GDD models typically assume that the GDD threshold for a species or even functional type is constant, but increasing evidence suggests otherwise. Climate can vary a lot on small spatial scales and recent studies suggest that fine-scale climate may matter to phenology. Here, we use observations from one urban arboretum and one rural forested site to assess GDDs until budburst using two methods for measuring climate data (i.e., weather station data and hobo logger data). We additionally incorporate simulations to better interpret our results and conclude with a series of simulated forecasts to estimate changes in these GDD model estimates. Our results suggest the urban arboretum site requires fewer GDDs until budburst and may have stronger microclimate effects than the rural forested site though these effects diminish with the use of hobo loggers. Additionally, we find that GDD models may become less accurate with warming and GDDs begin to accumulate at a faster rate. Our findings suggest we need to either use a method that is less reliant on accumulated, climatological sums or we must scrutinize results through the use of mixed models and simulated data as we demonstrate here.

Introduction

Understanding and predicting plant phenology in temperate deciduous forests is critical as it both shapes community structure and also influences major ecosystem services such as resource and forest management. Climate change and urbanization are advancing spring timing—such as budburst and leafout, which are strongly cued by temperature, resulting in longer growing seasons (Chuine *et al.*, 2001) which ultimately impacts these services. Spring budburst timing can have cascading effects to pollinators (Boggs & Inouye, 2012; Pardee *et al.*, 2017), on albedo (Williamson *et al.*, 2016), and on carbon dynamics (Richardson *et al.*, 2013). Temperate forests sequester carbon and help mitigate the negative effects of climate change and—with earlier spring phenology and longer growing seasons—there has been an increase in carbon uptake (Keenan *et al.*, 2014). Because of the importance of phenology, forecasting it accurately with climate change is a major and important aim across several fields of science (Moorcroft *et al.*, 2001; Taylor & White, 2020; Yu *et al.*, 2016; Zhao *et al.*, 2013).

One major forecasting method across all these fields is the growing degree day model. The growing degree day (GDD) model allows researchers to track heat accumulation to measure spring budburst (Cook *et al.*, 2012; Crimmins & Crimmins, 2019; Phillimore *et al.*, 2013; Schwartz *et al.*, 2006; Vitasse *et al.*, 2011). The

GDD model simply sums temperatures above a certain threshold—often around 0°C for forested trees as estimates are proven to be more accurate (Man & Lu, 2010)—and different species often require a different number of GDDs to leaf out. GDDs accumulate at a faster rate when mean temperatures are higher, thus different sites or different climate measurement methods may record different GDD thresholds for budburst. Integrating the GDD model successfully is essential for predicting the effects of climate change on systems where the climate is rapidly changing, including temperate forests.

We often assume in GDD models that the GDD required for a species, or even a suite of species (e.g., plant functional types) is constant, but increasing evidence suggests it may not be. The plasticity of phenology means that the same individual exposed to different climates will leafout at a very different time. Decades of work show that chilling—related to winter temperatures—and photoperiod can shift the GDD a plant needs for the same event (Basler & Körner, 2012; Chuine, 2010; Zohner *et al.*, 2016). Spring phenology also has a genetic component, and the required chilling, photoperiod and GDD can vary by population (Scotti-Saintagne *et al.*, 2004; Cuervo-Alarcon *et al.*, 2018). Though this genetic effect seems smaller than for other phenological events (McKown *et al.*, 2013; Satake *et al.*, 2013).

Climate, both on a larger or smaller scale, helps determine the role of chilling and photoperiod. On a large scale, there are climate gradients across space (i.e., latitudinal or continentality effects), but also gradients due to anthropogenic impact. Urbanization has led to the formation of urban heat islands, which have been shown to affect plant phenology and lead to earlier spring leafout (Meng *et al.*, 2020). Because urban sites strongly contribute to carbon sequestration (Ziter & Turner, 2018), these trends are crucial to understand in order to predict plant development with warming.

Increasingly, researchers have suggested that urban environments provide a natural laboratory for assessing the effects of warming on temperate tree and shrub species as these sites are warming at a faster rate than more rural habitats (Pickett *et al.*, 2011; Grimm *et al.*, 2008). Additionally urban sites often house arboreta or botanical gardens that often contribute long-term records (Zohner & Renner, 2014) or are used for experiments on phenology (Ettinger *et al.*, 2018). Arboreta and botanical gardens offer a unique lens to investigate climate change and local adaptation studies by incorporating varying seed sources—or provenance locations—thus they mimic common garden experiments (Primack & Miller-Rushing, 2009). It is essential that we have a better understanding if results from such urban sites directly translate to more natural forests.

Climate on a smaller scale may also be important to consider. Climate can vary significantly on small spatial scales (de Rességuier *et al.*, 2020; Lenoir *et al.*, 2013, e.g., as much as 2.6°C between sensors at the same vineyard or up to 6.6°C within 1 km spatial units in northern Europe). Thus, increasing evidence suggests that fine-scale climate may matter to phenology (Lembrechts *et al.*, 2019). Additionally, temperature variation at the bud level is shown to affect budburst timing within an individual canopy (Lembrechts *et al.*,

2019). To facilitate scaling and minimize error due to these fine-scal climatic effects (i.e., microclimate effects), researchers often deploy standalone weather loggers—such as HOBO sensors—which may provide higher resolution weather data (Schwartz *et al.*, 2013; Whiteman *et al.*, 01 Jan. 2000). Though deploying temperature loggers is not always feasible, especially when investigating large spatiotemporal shifts in GDDs.

Provenance location may also matter, though evidence is lacking for spring phenological phases (Aitken & Bemmels, 2015; McKown *et al.*, 2013). There is large debate over the role of provenance latitude on budburst initiation and the associated shifts in phenological cue use. Some studies suggest that: (1) species from lower latitudes will be more reliant on photoperiod with climate change (Zohner *et al.*, 2016), (2) photoperiod will slow or constrain range expansion (Saikkonen *et al.*, 2012), (3) all species will rely on photoperiod more as winters warm (Way & Montgomery, 2015), and (4) lower latitude species will require both strong photoperiod cues and more forcing in order to compensate for the lack of chilling but photosensitivity may be more important at the cold trailing edge for range expansion to occur (Gauzere *et al.*, 2017). Many arboreta keep diligent acquisition records, providing visitors and scientists information on seed sources (Dosmann, 2006), and the potential to test such provenance effects.

Here, we aimed to address the following hypotheses: (1) required GDD in an urban arboreta will vary from a rural forested site. We predicted lower chilling in the urban site could lead to greater required GDD, (2) microclimates will lead to variation in GDD within sites, (3) individuals with provenance latitudes from more northern locations require fewer GDDs to budburst. We tested these in one urban arboretum and one rural forested site and we additionally incorporated simulations to help better interpret our results.

Methods

Sites

We chose two sites—one urban arboretum and one forest—with overlapping species and climates to compare the number of growing degree days to budburst across species. The urban site is in Boston, MA at the Arnold Arboretum of Harvard University (42°17' N -71°8' W). The Arnold Arboretum is 281 acres, contains 3825 woody plant taxa from North America, Europe and Asia and has an elevation gain of approximately 96m. The forest site is in Petersham, MA at the Harvard Forest (42°31'53.5' N -72°11'24.1' W). The Harvard Forest is 1446 acres and has a range of elevation of 220-410m.

Simulations

We simulated test data in order to assess our model output results, especially our inference on teasing out effects of microclimates versus provenance versus potential differences across weather station and hobo logger data. Our simulations were designed to test the following potential effects: (1) urban environments require more GDDs, (2) presence of microclimates (at one or both sites) accurately measured by hobo loggers, (3) presence of provenance effects (i.e., there were multiple provenance latitudes at the urban arboretum site but only one at the rural forest site), (4) weather stations or hobo loggers are effectively 'noisier' data for GDD models compared to the other.

To run our simulations, we assumed each species needed a different GDD (drawing each species' requirement from a normal distribution). We then modeled climate data by again establishing a random distribution around a mean temperature for each site. Using this climate data, we found the day of budburst when the unique GDD threshold was met for each individual. To test that urban sites require more GDD, we created simulation data that manipulates the GDD threshold for the urban versus rural sites by increasing the GDD threshold for individuals at the more urban locations (e.g., local arboreta). To test microclimate effects, we built our climate data then add variation to this weather data to create "microclimate" effects. To test the provenance latitude hypothesis we make individuals from more northern provenances require fewer GDDs. To test for the effect of noise, we add noise by increasing the standard deviation value for our random distribution around a mean temperature for each method.

We additionally examined the accuracy of GDD models using different base temperature thresholds in combination with warming through simulations. To evaluate the accuracy of GDD models, we used different base temperatures for calculating GDD (i.e., $0°C$ versus $10°C$) with variation in sigma around base temperatures (i.e., $0.1°C$ and $0.5°C$). We also tested GDD accuracy across various GDD threshold requirements with warming of $1°C$ to $10°C$ and using varying GDD threshold requirements for budburst. Accuracy was evaluated as a ratio of observed GDD divided by the expected GDD.

Data analysis

Using Bayesian hierarchical models with the rstan package (Stan Development Team, 2019), version 2.19.2, in R (R Development Core Team, 2017), version 3.3.1, we estimated the effects of urban or provenance effect and method effect and all two-way interactions as predictors on GDDs until budburst. Species were modeled hierarchically as grouping factors, which generates an estimate and posterior distribution of the overall response across the 15 species used in our simulations and 18 species used in our real data. We ran four chains, each with 2 500 warm-up iterations and 3 000 iterations for a total of 2 000 posterior samples for

each predictor for each model using weakly informative priors. Increasing priors three-fold did not impact our results. We evaluated our model performance based on $\hat{R}$ values that were close to one and did not include models with divergent transitions in our results. We also evaluated high n_{eff} (2000 for most parameters, but as low as 708 for a couple of parameters in the simulated provenance latitude model). We additionally assessed chain convergence and posterior predictive checks visually (Gelman *et al.*, 2014). We report means ± 50% uncertainty intervals from our models in the main text because these intervals are more computationally stable (Carpenter *et al.*, 2017; Gelman *et al.*, 2014). See Tables S1-S7 for 95% uncertainty intervals. In model output figures, we also report explained variance (i.e., the 'sigma' values) around key parameters from the model output, which are important to consider for understanding partitioning of error within the model (Gelman *et al.*, 2014).

Shiny App

To show the above simulations, real data and forecasts in one location we use a Shiny Application. Using the R package 'shiny' (Chang *et al.*, 2021), version 1.6.0, we developed a Shiny App that contains five pages: (1) 'Home' which has information on the application, (2) 'Hypothesis Testing' which runs the simulation data and allows users to manipulate the inputs, (3) 'Simulation Data for Model Testing' which runs simulation data to test the model and make sure the model outputs are accurate, (4) 'Real Data and Analyze Results' which uses real data and runs analyses to be used to compare to the 'Hypothesis Testing' output and (5) 'Forecasting GDD with Warming' which forecasts GDD accuracy under warming.

Results

Simulations

We find we can accurately recover a simple effect of urban sites requiring more GDDs until budburst (Figure 2 **a)** and Table S1). Simulations that include microclimates at both sites show that the hobo loggers require more GDDs until budburst. When simulating microclimate effects—thus greater variation in GDD—across the sites, we include greater variation in temperature for the hobo logger data, which is being reflected by the negative slope of the method parameter (Figure 2 **b)** and Table S2). Provenance effect simulations indicate more northern provenance locations require fewer GDDs until budburst which is being recovered in the provenance parameter (Figure 2 **c)** and Table S3). When we manipulate the simulations to have noisy weather station data, noise is returned as the sigma for the method parameter (Figure 2 **d)** and Table S4) and the method parameter is slightly positive, indicating weather stations require more GDDs until budburst.

Though, when we manipulate the simulations to have noisy hobo logger data, the output is nearly identical (Figure 2 **b**) and Table S5) b~~ut the method parameter is slightly negative,~~ now indicating hobo loggers require more GDDs until budburst.

Simulations: GDD accuracy

The GDD model becomes less accurate with warming, and accuracy decreases at a faster rate with the higher base temperature (i.e., 10°C) than with the lower base temperature (i.e., 0°C; Figure 5). Under the no warming simulation, using the 10°C base temperature is most consistent across species but with any amount of warming, the 0°C base temperature is more accurate. The GDD model is most accurate for individuals that have high GDD thresholds and when base temperatures are higher (i.e., 10°C; Figure S5). Additionally, variability in accuracy across GDD thresholds and with warming increases with higher sigmas (Figure S5 and Figure 5).

Empirical data

Mean spring temperature at the urban arboretum site was 4.39°C and was 1.42°C at the rural forested site using weather station climate data (Figure 1). When climate data was recorded with the hobo loggers, mean spring temperature at the urban arboretum was 6.13 and was 1.78°C at the rural forested site (Figure 1). Overall, the hobo loggers generally recorded higher temperatures than the weather station at the urban arboretum site (with a mean of 1.75°C and a standard deviation of 1.03°C; Figure S2). The rural forested sight generally had more variation around the weather station, though did not typically record higher or lower temperatures than the weather station: the mean difference was 0.55°C with a standard deviation of 1.04°C (Figure S2).

Individuals at the urban arboretum site require fewer GDDs to budburst (-32.57 ± 15.94) than the individuals at the rural forested site. We also found there is high variation in GDDs between the two methods (16.72) though the slope is close to zero (method.real ± method.real.sd). There is a large interaction indicating weather station data at the arboretum records the fewest number of GDDs until budburst (-40.29 ± 16.18). In fact, the interactive effect was the strongest predictor of GDDs, even stronger than the effect of site. By disentangling the recorded temperatures between the two methods across the two sites, we see that there is greater variation in hobo logger temperatures at the urban arboretum and these temperatures are generally recording higher temperatures than the weather station (Figure 1 and Figure S2). Using raw data combined with model output, we see ~~there is~~ higher variation at the arboretum across the two methods and that the arboretum requires fewer GDDs until budburst than the rural forested site (Figure 4). We also see that

the hobo loggers across the two sites report similar estimates of GDDs until budburst, whereas the weather station at the arboretum is reporting much lower GDDs until budburst than the rural forest weather station (Figure 4).

When testing for provenance effects, there was no overall effect of provenance latitude on GDDs until budburst and there was high variation (18.37 ± 11.87). There was still little effect of method on GDDs until budburst and large variation (-8.52 ± 8.18). The interaction of provenance by method was close to zero (-3.53 ± 12.49) but the sigma was large (13.88).

Discussion

Our study assessed the effects of an urban arboretum versus a more rural forested site coupled with the effect of climate recording method (i.e., weather station versus hobo logger temperature data) on GDD until budburst. We found the urban site was in fact warmer, but this did not translate to individuals requiring more GDDs but rather fewer GDDs until budburst. Our results additionally suggest there was a strong microclimate effect as is apparent by the large variation in GDD with method. Though these effects varied by site, with hobo loggers at the urban arboretum generally recording higher temperatures than the weather station but with hobo loggers at the rural forested site recording more variation in temperatures than the associated weather station. Provenance did not determine clear results so we suggest teasing out provenance effects, given that they may not contribute much to spring phenology (Gauzere *et al.*, 2017) and it is difficult with our paucity of latitudinal variation and low sample size for various provenances.

Variation across and within sites suggests important variation for forecasting

Our finding that urban trees require fewer GDDs contributes to increasing evidence that trees in urban sites may respond differently than those in forested rural areas. Additionally, our results are in line with other recent studies investigating rural versus urban sites with a lower GDD requirement at urban sites and thus a lower temperature sensitivity (Meng *et al.*, 2020) at these sites, with colder rural sites requiring even more GDDs until budburst than their associated urban site. This means that long-term records and experiments conducted in arboreta in urban areas may not be transferable to larger scale studies and models that incorporate forested rural areas. We should perhaps be more cautious in these extrapolations. However, we see that the urban effect is weaker when we used hobo loggers at both sites. Further studies that investigate more rural and associated urban sites are necessary to test if hobo loggers can lessen the urban effect we are seeing here.

The lower GDD requirement until budburst is likely due to the arboretum accumulating more over-winter chilling, however, our results become more similar when we use the hobo loggers to measure GDD. Recent research suggests individuals can accumulate chilling at temperatures as high as 10°C (Baumgarten *et al.*, 2021)—or even up to 15°C in subtropical trees (Zhang *et al.*, 2021)—but the duration of winter is more important than the temperature. If average temperatures are below the chilling accumulation threshold, which may occur at cooler sites, then we can expect less over-winter chilling accumulation at colder sites. These results suggests we use caution when using urban sites as natural experiments as these sites may not mirror forest habitats, especially when sites are from colder (e.g., more northern) regions.

When looking at both sites, we found no consistent differences in GDDs estimated between the weather stations and the hobo loggers, but there was a large interaction, which suggests large microclimatic differences across the two sites or within one site but not the other. At the rural forested site, we see that there is indeed greater variation in temperatures recorded from the hobo loggers than the weather station but that variation is lower than at the urban arboretum. The urban arboretum weather station tends to record cooler temperatures than most of the hobo loggers (Figure 1) and since these differences in temperature occur close to 0°C (with some hobo loggers contributing to accumulated GDDs but the weather station under the 0°C threshold), these differences could in fact be biologically meaningful to phenology. In contrast, the climate data from the hobo logger and the weather station seem more similar at the rural forested site. And since GDDs are predominant indicators of spring phenology, having accurate and consistent weather data is essential for better estimates of budburst or leafout, especially with warming. By incorporating both hobo logger and weather station data, we are able to determine microclimate effects across the two sites and detect nuances in canopy temperature variation across two different types of sites: generally open-canopy, urban arboretum versus a typically closed-canopy, rural forest. Given these results, we may see a stronger microclimate effect at the urban arboretum than the rural forested site—which could be true given different tree species create different under canopy effects, or there are more sidewalks and roads, etc.—though more work is necessary to be certain.

Accurately attributing observed variation requires more research on climate methods and phenology

Determining which methods are most accurate is the first step to determining fine-scale climatic variation and, ultimately, better forecast phenology under climate change. Our simulations suggest teasing out noise versus microclimate effects could be difficult. As is evident from the model outputs, the method which is less accurate requires more GDDs until budburst since GDDs are accumulating at a faster rate with higher temperature variability—and higher temperature days. By including microclimate effects at both sites,

variation in temperature increases and, thus, the number of days in which the temperature falls below the base GDD threshold is also greater but the days in which temperatures are accumulating, the temperature will likely be greater than what is recorded at the weather station. High temperature variability, whether it is due to inaccurate recordings or due to microclimates, results in more days at higher temperatures so the day of budburst records higher GDDs. As climate is one of the strongest environmental factors contributing to ecosystem change, it is essential to measure weather data as accurately and efficiently as possible, while using methods that are accessible to myriad researchers.

Future studies that investigate local climate and phenology are essential and here we suggest several pathways to more accurately model GDDs under climate change. Using field studies, we suggest the implementation of more intensive climate method research including specific studies of hobo loggers and the effects of radiation shields and location in the canopy and to apply these treatments next to both weather stations and to the trees or shrubs of interest. We also need a better understanding of what temperatures (e.g., bud temperature versus air temperature) are actually important to plant phenology, especially studies that tease out radiative heating effects. And finally, we propose the use of better models as we found the use of just GDD values may not be as good as models that incorporate daily climate data or, better still, using process-based models (Keenan *et al.*, 2020).

As we see from our forecasting simulations, regardless of base temperature threshold, GDD models may not be appropriate for the future with warming (Man & Lu, 2010). This is because with warming, GDDs will accumulate at a faster rate, which will reduce accuracy of determining that actual threshold for budburst phenology. Generally higher GDD thresholds means a lower GDD observed to GDD expected ratio; this is because being off by a day is a small effect for higher GDD threshold species (and hence greater days) than for lower GDD threshold species, but it really depends on climate variability because high variability means some days you can get accumulate GDDs quickly and that can override the GDD threshold trends. In reality temperature variance likely changes over the spring, rendering climate change effects even harder to decipher. In the future, we need to either use a method that is less reliant on accumulated sums—especially if it is a climatological sum—or we must scrutinize results through the use of mixed models and simulated data as we demonstrate here.

Acknowledgments

O'Keefe for his work and observations at the Harvard Forest. We also want to thank W. Daly, D. Buonaiuto, M. Garner, J. Gersony, F. Jones, G. Legault, D. Loughnan, A. Manandhar, A. O'Regan and D. Sodhi for their continued feedback and insights that helped improved the experimental design, models, simulations and manuscript.

Author Contribution

C.J.C. and E.M.W. conceived of the study, identified hypotheses to test in the study and determined which sites to observe. C.J.C. performed the analyses and produced all figures and tables. C.J.C. wrote the paper, and both authors edited it.

Data Availability

Data and code from the analyses will be available via the Harvard Forest Data Archive upon publication. Raw data, Stan model code and output are available on GitHub and provided upon request.

References

Aitken, S.N. & Bemmels, J.B. (2015) Time to get moving: assisted gene flow of forest trees. *Evolutionary Applications* **9**, 271–290.

Basler, D. & Körner, C. (2012) Photoperiod sensitivity of bud burst in 14 temperate forest tree species. *Agricultural and Forest Meteorology* **165**, 73–81.

Baumgarten, F., Zohner, C.M., Gessler, A. & Vitasse, Y. (2021) Chilled to be forced: the best dose to wake up buds from winter dormancy. *New Phytologist* **230**, 1366–1377.

Boggs, C.L. & Inouye, D.W. (2012) A single climate driver has direct and indirect effects on insect population dynamics. *Ecology Letters* **15**, 502–508.

Carpenter, B., Gelman, A., Hoffman, M., Lee, D., Goodrich, B., Betancourt, M., Brubaker, M.A., Guo, J., Li, P. & Allen, R. (2017) Stan: A probabilistic programming language. *Journal of Statistical Software* **76**, 10.18637/jss.v076.i01.

Chang, W., Cheng, J., Allaire, J., Sievert, C., Schloerke, B., Xie, Y., Allen, J., McPherson, J., Dipert, A. & Borges, B. (2021) shiny: Web application framework for r. https://CRAN.R-project.org/package=shiny.

Chuine, I. (2010) Why does phenology drive species distribution? *Philosophical Transactions of the Royal Society B: Biological Sciences* **365**, 3149–3160.

Chuine, I., Aitken, S.N. & Ying, C.C. (2001) Temperature thresholds of shoot elongation in provenances of *Pinus contorta*. *Canadian Journal of Forest Research* **31**, 1444–1455.

Cook, B.I., Wolkovich, E.M., Davies, T.J., Ault, T.R., Betancourt, J.L., Allen, J.M., Bolmgren, K., Cleland, E.E., Crimmins, T.M., Kraft, N.J.B. & et al. (2012) Sensitivity of spring phenology to warming across temporal and spatial climate gradients in two independent databases. *Ecosystems* **15**, 1283–1294.

Crimmins, M.A. & Crimmins, T.M. (2019) Does an early spring indicate an early summer? relationships between intraseasonal growing degree day thresholds. *Journal of Geophysical Research: Biogeosciences* **124**, 2628–2641.

Cuervo-Alarcon, L., Arend, M., Müller, M., Sperisen, C., Finkeldey, R. & Krutovsky, K.V. (2018) Genetic variation and signatures of natural selection in populations of european beech (fagus sylvatica l.) along precipitation gradients. *Tree Genetics & Genomes* **14**, 84.

de Rességuier, L., Mary, S., Le Roux, R., Petitjean, T., Quénol, H. & van Leeuwen, C. (2020) Temperature variability at local scale in the bordeaux area. relations with environmental factors and impact on vine phenology. *Frontiers in Plant Science* **11**, 515.

Dosmann, M.S. (2006) Research in the garden: Averting the collections crisis. *The Botanical Review* **72**, 207–234.

Ettinger, A.K., Gee, S. & Wolkovich, E.M. (2018) Phenological sequences: how early-season events define those that follow. *Am J Bot* **105**, 1771–1780.

Gauzere, J., Delzon, S., Davi, H., Bonhomme, M., Garcia de Cortazar-Atauri, I. & Chuine, I. (2017) Integrating interactive effects of chilling and photoperiod in phenological process-based models. A case study with two European tree species: *Fagus sylvatica* and *Quercus petraea*. *Agricultural and Forest Meteorology* **244-255**, 9–20.

Gelman, A., Carlin, J.B., Stern, H.S., Dunson, D.B., Vehtari, A. & Rubin, D.B. (2014) *Bayesian Data Analysis*. CRC Press, New York, 3rd edn.

Grimm, N.B., Faeth, S.H., Golubiewski, N.E., Redman, C.L., Wu, J., Bai, X. & Briggs, J.M. (2008) Global change and the ecology of cities. *Science* **319**, 756–760.

Keenan, T.F., Gray, J., Friedl, M.A., Toomey, M., Bohrer, G., Hollinger, D.Y., Munger, J.W., O'Keefe, J., Schmid, H.P., Wing, I.S. & et al. (2014) Net carbon uptake has increased through warming-induced changes in temperate forest phenology. *Nature Climate Change* **4**, 598–604.

Keenan, T.F., Richardson, A.D. & Hufkens, K. (2020) On quantifying the apparent temperature sensitivity of plant phenology. *New Phytologist* **225**, 1033–1040.

Lembrechts, J.J., Nijs, I. & Lenoir, J. (2019) Incorporating microclimate into species distribution models. *Ecography* **42**, 1267–1279.

Lenoir, J., Graae, B.J., Aarrestad, P.A., Alsos, I.G., Armbruster, W.S., Austrheim, G., Bergendorff, C., Birks, H.J.B., Bråthen, K.A., Brunet, J., Bruun, H.H., Dahlberg, C.J., Decocq, G., Diekmann, M., Dynesius, M., Ejrnæs, R., Grytnes, J.A., Hylander, K., Klanderud, K., Luoto, M., Milbau, A., Moora, M., Nygaard, B., Odland, A., Ravolainen, V.T., Reinhardt, S., Sandvik, S.M., Schei, F.H., Speed, J.D.M., Tveraabak, L.U., Vandvik, V., Velle, L.G., Virtanen, R., Zobel, M. & Svenning, J.C. (2013) Local temperatures inferred from plant communities suggest strong spatial buffering of climate warming across northern europe. *Global Change Biology* **19**, 1470–1481.

Man, R. & Lu, P. (2010) Effects of thermal model and base temperature on estimates of thermal time to bud break in white spruce seedlings. *Canadian Journal of Forest Research* **40**, 1815–1820.

McKown, A.D., Guy, R.D., Klápště, J., Geraldes, A., Friedmann, M., Cronk, Q.C.B., El-Kassaby, Y.A., Mansfield, S.D. & Douglas, C.J. (2013) Geographical and environmental gradients shape phenotypic trait variation and genetic structure inpopulus trichocarpa. *New Phytologist* **201**, 1263–1276.

Meng, L., Mao, J., Zhou, Y., Richardson, A.D., Lee, X., Thornton, P.E., Ricciuto, D.M., Li, X., Dai, Y., Shi, X. & et al. (2020) Urban warming advances spring phenology but reduces the response of phenology to temperature in the conterminous united states. *Proceedings of the National Academy of Sciences* **117**, 4228–4233.

Moorcroft, P.R., Hurtt, G.C. & Pacala, S.W. (2001) A method for scaling vegetation dynamics: The Ecosystem Demography Model (ED). *Ecological Monographs* **71**, 557–585.

Pardee, G.L., Inouye, D.W. & Irwin, R.E. (2017) Direct and indirect effects of episodic frost on plant growth and reproduction in subalpine wildflowers. *Global Change Biology* **24**, 848–857.

Phillimore, A.B., Proios, K., O'Mahony, N., Bernard, R., Lord, A.M., Atkinson, S. & Smithers, R.J. (2013) Inferring local processes from macroscale phenological pattern: a comparison of two methods. *Journal of Ecology* **101**, 774–783.

Pickett, S.T., Cadenasso, M.L., Grove, J.M., Boone, C.G., Groffman, P.M., Irwin, E., Kaushal, S.S., Marshall, V., McGrath, B.P., Nilon, C.H. *et al.* (2011) Urban ecological systems: Scientific foundations and a decade of progress. *Journal of environmental management* **92**, 331–362.

Primack, R.B. & Miller-Rushing, A.J. (2009) The role of botanical gardens in climate change research. *New Phytologist* **182**, 303–313.

R Development Core Team (2017) R: A language and environment for statistical computing. *R Foundation for Statistical Computing, Vienna, Austria* .

Richardson, A.D., Keenan, T.F., 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.

Saikkonen, K., Taulavuori, K., Hyvönen, T., Gundel, P.E., Hamilton, C.E., Vänninen, I., Nissinen, A. & Helander, M. (2012) Climate change-driven species' range shifts filtered by photoperiodism. *Nature Climate Change* **2**, 239.

Satake, A., Kawagoe, T., Saburi, Y., Chiba, Y., Sakurai, G. & Kudoh, H. (2013) Forecasting flowering phenology under climate warming by modelling the regulatory dynamics of flowering-time genes. *Nature Communications* **4**, 2303.

Schwartz, M.D., Ahas, R. & Aasa, A. (2006) Onset of spring starting earlier across the Northern Hemisphere. *Global Change Biology* **12**, 343–351.

Schwartz, M.D., Hanes, J.M. & Liang, L. (2013) Comparing carbon flux and high-resolution spring phenological measurements in a northern mixed forest. *Agricultural and Forest Meteorology* **169**, 136–147.

Scotti-Saintagne, C., Bodénès, C., Barreneche, T., Bertocchi, E., Plomion, C. & Kremer, A. (2004) Detection of quantitative trait loci controlling bud burst and height growth in quercus robur l. *Theoretical and Applied Genetics* **109**, 1648–1659.

Stan Development Team (2019) Stan user's guide, version 2.19.

Taylor, S.D. & White, E.P. (2020) Automated data-intensive forecasting of plant phenology throughout the united states. *Ecological Applications* **30**, e02025.

Vitasse, Y., Francois, C., Delpierre, N., Dufrene, E., Kremer, A., Chuine, I. & Delzon, S. (2011) Assessing the effects of climate change on the phenology of European temperate trees. *Agricultural and Forest Meteorology* **151**, 969–980.

Way, D.A. & Montgomery, R.A. (2015) Photoperiod constraints on tree phenology, performance and migration in a warming world. *Plant, Cell & Environment* **38**, 1725–1736.

Whiteman, C.D., Hubbe, J.M. & Shaw, W.J. (01 Jan. 2000) Evaluation of an inexpensive temperature datalogger for meteorological applications. *Journal of Atmospheric and Oceanic Technology* **17**, 77 – 81.

Williamson, S.N., Barrio, I.C., Hik, D.S. & Gamon, J.A. (2016) Phenology and species determine growing-season albedo increase at the altitudinal limit of shrub growth in the sub-arctic. *Global Change Biology* **22**, 3621–3631.

Yu, M., Wang, G. & Chen, H. (2016) Quantifying the impacts of land surface schemes and dynamic vegetation on the model dependency of projected changes in surface energy and water budgets. *Journal of Advances in Modeling Earth Systems* **8**, 370–386.

Zhang, R., Lin, J., Wang, F., Shen, S., Wang, X., Rao, Y., Wu, J. & Hänninen, H. (2021) The chilling requirement of subtropical trees is fulfilled by high temperatures: A generalized hypothesis for tree endodormancy release and a method for testing it. *Agricultural and Forest Meteorology* **298-299**, 108296.

Zhao, M., Peng, C., Xiang, W., Deng, X., Tian, D., Zhou, X., Yu, G., He, H. & Zhao, Z. (2013) Plant phenological modeling and its application in global climate change research: overview and future challenges. *Environmental Reviews* **21**, 1–14.

Ziter, C. & Turner, M.G. (2018) Current and historical land use influence soil-based ecosystem services in an urban landscape. *Ecological Applications* **28**, 643–654.

Zohner, C.M., Benito, B.M., Svenning, J.C. & Renner, S.S. (2016) Day length unlikely to constrain climate-driven shifts in leaf-out times of northern woody plants. *Nature Climate Change* **6**, 1120–1123.

Zohner, C.M. & Renner, S.S. (2014) Common garden comparison of the leaf-out phenology of woody species from different native climates, combined with herbarium records, forecasts long-term change. *Ecology Letters* **17**, 1016–1025.
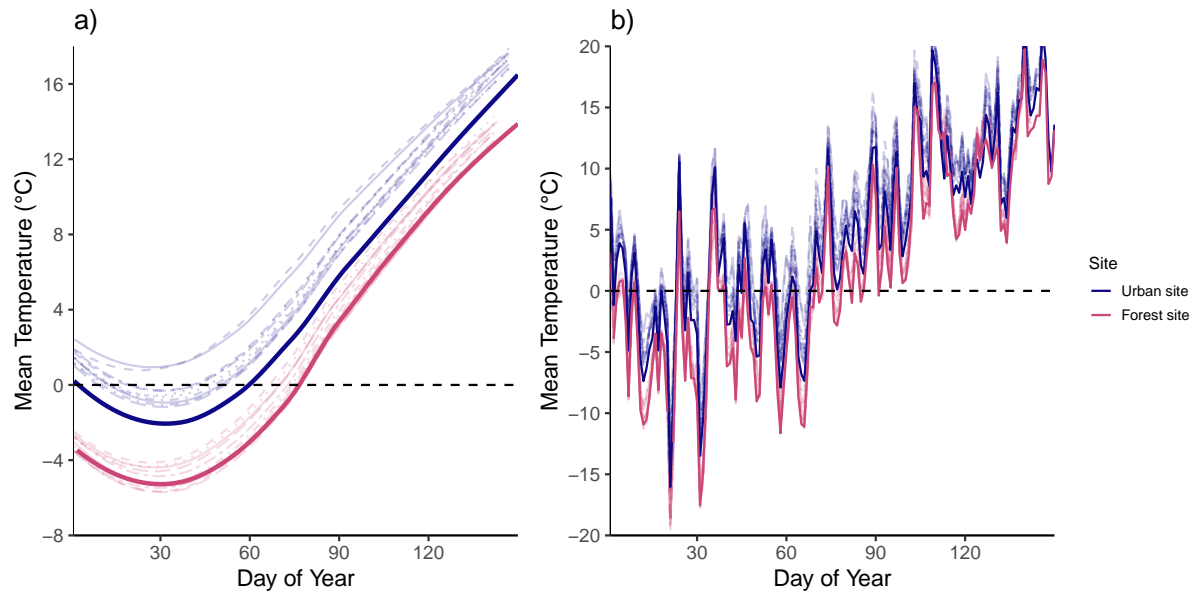
Tables and Figures

Figure 1: Here we show a breakdown of the climate data across the two sites with darker lines representing weather station data and the lighter, more transparent lines of varying line types representing the hobo loggers: a) a series of smoothing splines of mean temperature with 90% credible interval and b) actual mean temperature.
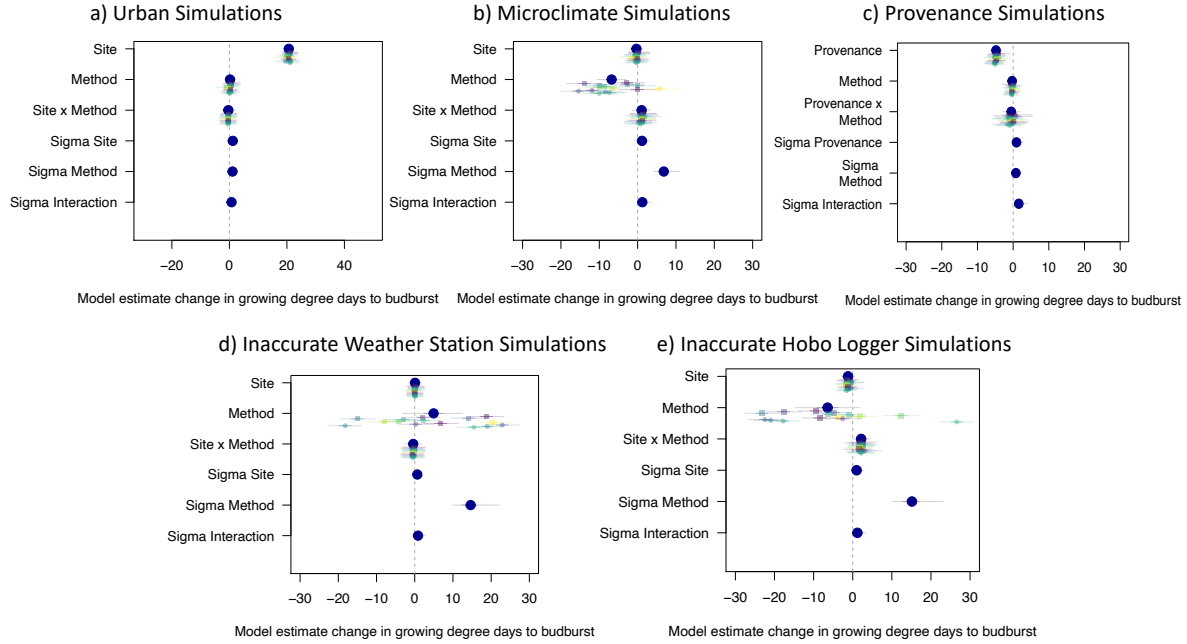
Figure 2: Simulations: using simulations we show (a) urban sites requiring more GDDs, (b) microclimate effects, (c) more northern provenance latitudes requiring fewer GDDs, (d) less accurate weather station data and (e) less accurate hobo logger data. We show the effects of site in (a), (b), (d) and (e) as a binary 'urban' parameter with the urban site as '1' and the rural site as '0'. We also use a binary 'weather station' parameter to represent the climate data method with weather station data reported as '1' and with hobo logger data as '0'. The intercept represents the hobo logger data for the rural forested site. More positive values indicate more GDDs are required for budburst whereas more negative values suggest fewer GDDs are required. Dots and thin lines show means and 90% uncertainty intervals and thick lines show 50% uncertainty intervals. See Tables S1, S2, S3, S4 and S5 for full model output.
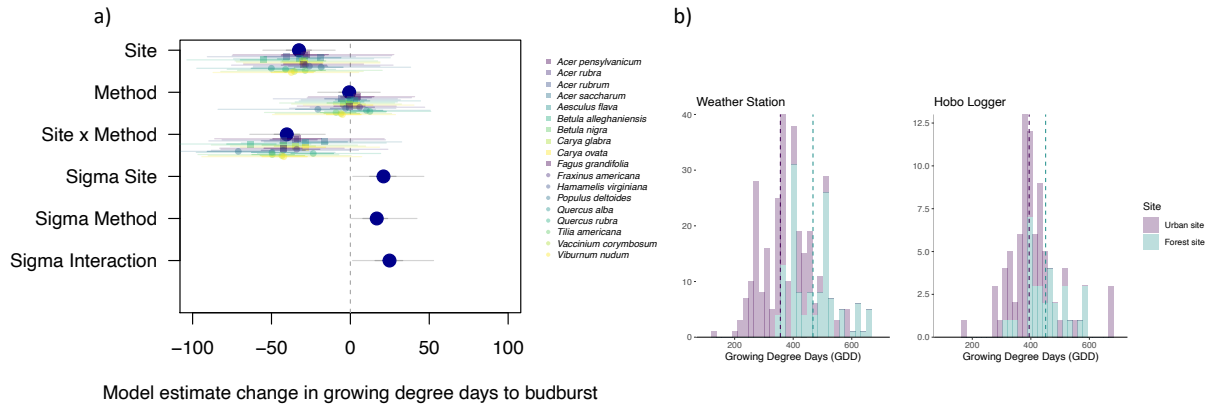
17

Figure 3: Empirical Data: Using real data, we show (a) the effects of site (again using a binary predictor with the urban arboretum site reported as '1' and the rural forested site reported as '0') and climate data method (as a binary predictor with weather station data as '1' and hobo logger data as '0') on growing degree days (GDDs) until budburst. The intercept represents the hobo logger data for the rural forested site. More positive values indicate more GDDs are required for budburst whereas more negative values suggest fewer GDDs are required. Dots and thin lines show means and 90% uncertainty intervals and thick lines show 50% uncertainty intervals. See Table S6 for full model output. We also show (b) histograms of GDDs at the urban arboretum and rural forested site using weather station data and hobo logger data.
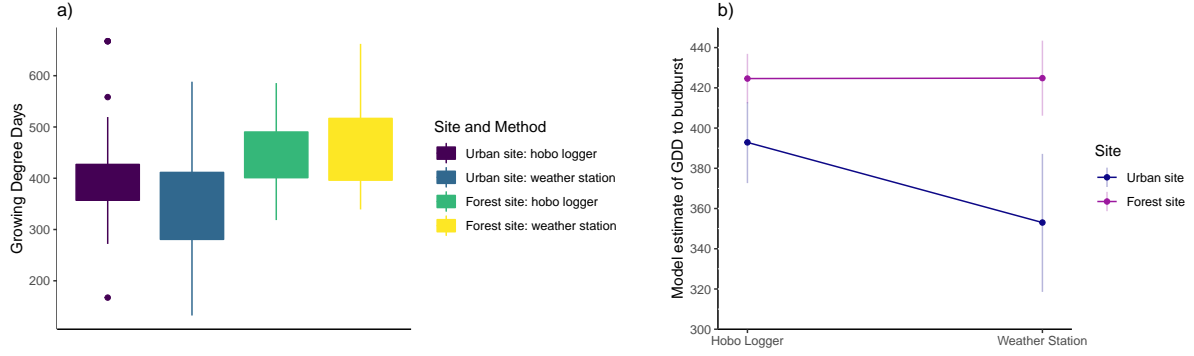
Figure 4: We show effects of site (urban arboretum site versus forested rural site) by climate data method (weather station data versus hobo logger data) on growing degree days (GDDs) until budburst (a) as a boxplot across each method and site combination using raw data and (b) using model output to show the mean estimates for each site and method with 50% uncertainty intervals shown as errorbars.
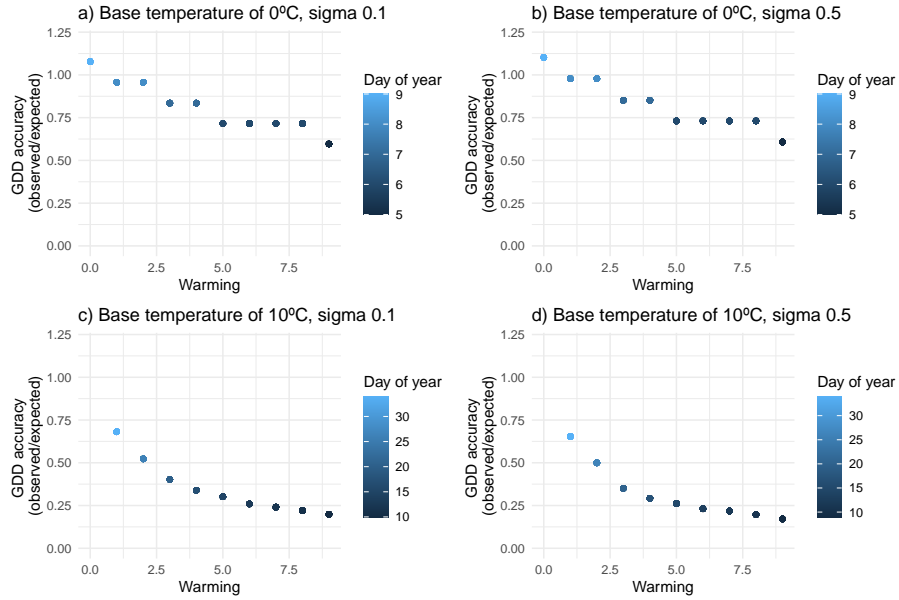

Figure 5: Using simulated data, we show how GDD measurement accuracy changes with warming (i.e., from 0°C to 10°C) using a base temperature of (a) 0°C and a sigma of 0.1°C, (b) 0°C and a sigma of 0.5°C, (c) 10°C and a sigma of 0.1°C and (d) 10°C and a sigma of 0.5°C. GDD accuracy is measured as the observed GDD divided by the expected GDD.