

Variation across space, species and methods in models of spring phenology

Authors:

C. J. Chamberlain^{1,2,3} & E. M. Wolkovich^{1,2,4}

Author affiliations:

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

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

³Conservation International, Arlington, VA, USA;

⁴Forest & Conservation Sciences, Faculty of Forestry, University of British Columbia, 2424 Main Mall, Vancouver, BC V6T 1Z4;

*Corresponding author: 248.953.0189; cchamberlain@conservation.org

Keywords: phenology, climate change, forest communities, microclimate, urban heat island, growing degree days

Highlights

- Individual trees growing in more urban settings require fewer GDDs to budburst.
- Understanding microclimatic effects across sites is essential for phenological forecasts.
- The GDD model varies in accuracy across species, location, and climate.
- Inaccuracy increases for early-active species using the GDD model.

Abstract

Predicting spring plant phenology in temperate forests is critical for forecasting important processes such as carbon storage, especially as climate change and urbanization shift many phenological phases. One major forecasting method for phenology is the growing degree day (GDD) model, which tracks heat accumulation. Forecasts using GDD models typically assume that the GDD threshold for a species, or even functional type, is constant across diverse landscapes, but increasing evidence suggests otherwise. Shifts in climate over time with anthropogenic warming, especially warmer winters, may alter the required GDD. Variation in climate across space may also lead to variation in GDD requirements, with recent studies suggesting that even fine-scale spatial variation in climate may matter to phenology. Here, we combine simulations, observations from an urban arboretum and a rural forested site, and Bayesian hierarchical models to assess how consistent GDD models of budburst are across species and space. We build GDD models using two different methods to measure climate data: weather station data and HOBO logger data. We find that estimated GDD thresholds can vary up to 20% across sites and methods. 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 depend on the method to measure climate. Further, we find that GDD models may become less accurate with warming, as GDDs begin to accumulate faster, and are less accurate for early-active species. Our results suggest that forecasts based on GDD models for spring phenology should incorporate these inherent accuracy issues of GDD models, and also the variations we found across space,

species and warming. Improved understanding of GDD models, alongside more mechanistic models and a more refined understanding of spring phenology, could improve forecasts for temperate forests.

Introduction

Understanding and predicting spring plant phenology in temperate deciduous forests is critical as it both shapes community structure and 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). These shifts in growing seasons ultimately impact ecosystem services.

Spring budburst timing in particular can have cascading effects on pollinators (Boggs & Inouye, 2012; Pardee *et al.*, 2017), albedo (Williamson *et al.*, 2016), and carbon dynamics (Richardson *et al.*, 2013). Temperate forests sequester carbon and help mitigate the negative effects of climate change; with earlier spring phenology and longer growing seasons forests have increased carbon uptake (Keenan *et al.*, 2014). Because of this importance, forecasting phenology accurately with climate change is a major and important aim across several fields of science including agronomy, ecology, evolution and hydrology (Moorcroft *et al.*, 2001; Bolton & Friedl, 2013; Yu *et al.*, 2016; Taylor & White, 2020).

One major forecasting method widely accepted and utilized across these fields is the growing degree day (GDD) model, which allows researchers to track heat accumulation to predict spring budburst (Schwartz *et al.*, 2006; Vitasse *et al.*, 2011; Cook *et al.*, 2012; Phillimore *et al.*, 2013; Crimmins & Crimmins, 2019). Studies suggest other models—particularly the Crop Heat Unit (Kumudini *et al.*, 2014; Abendroth *et al.*, 2019)—are more precise than the GDD model, though the GDD model is more widely accepted for tree and shrub species (Yu *et al.*, 2016; Klosterman *et al.*, 2018; Montgomery *et al.*, 2020). The model takes the mean daily temperature, subtracts this value from the threshold temperature—often 0°C for forest trees (as estimates are proven to be more accurate, Man & Lu, 2010)—and sums these temperatures each day. Different species

generally require a different number of GDDs to leaf out, with early-leafout species requiring less than later-leafout species. 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. Understanding the complexity of the apparently simple GDD model (Bonhomme, 2000) is essential for predicting the effects of climate change on systems where the climate is rapidly changing, including temperate forests.

Forecasts using GDD models often assume 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).

Climate helps determine the role of chilling and photoperiod—and, thus the required GDD. 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 can 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 important to understand to best 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 warm at a faster rate than more rural habitats (Grimm *et al.*, 2008; Pickett *et al.*, 2011). Additionally urban sites often house arboreta or botanical gardens that often contribute long-term phenology 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 (i.e., origin) locations—thus mimicking common garden experiments (Primack & Miller-Rushing, 2009). Given these important roles of urban sites and the arboreta within them, understanding if results from urban sites directly translate to more natural forests has implications for both basic science and forecasts.

Climate on a smaller scale may also be important to determining spring phenology as it can vary significantly (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, Lenoir *et al.*, 2013; de Ressaiguier *et al.*, 2020). Increasing evidence suggests that fine-scale climate may matter to phenology (Lembrechts *et al.*, 2019). To facilitate scaling and minimize error due to these fine-scale climatic effects, which we refer to as 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).

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 fall phenological events (McKown *et al.*, 2013; Aitken & Bemmels, 2015; Vico *et al.*, 2021). Currently, there is debate over the directional effect of provenance latitude on budburst timing 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 (rather than the warm) range edge (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 less chilling accumulation in the urban site could lead to greater required GDD. (2) Individuals from more northern provenance locations will require fewer GDDs to budburst, and (3) microclimate effects will lead to variation in GDD within sites. We tested these in an urban arboretum, which included a diversity of provenances, and a rural forested site, and incorporated simulations to better interpret our results.

Methods

Sites & Species

We chose two sites—one urban arboretum and one rural 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 13-73 m. We used budburst observations (i.e., defined as the ‘beginning of sprouting or bud breaking; shoot emergence’ as BBCH scale 07 Finn *et al.*, 2007) from 77 individuals collected by citizen science volunteers from the TreeSpotters group at the Arnold Arboretum, facilitated through the National Phenology Network (<https://www.usanpn.org/taxonomy/term/438>). Observations were typically recorded every one to two days by at least one volunteer.

Of the total tree individuals observed, 26 had provenance latitude information ranging from 33.79 to 52.54, whereas the remaining 51 individuals originated at the Arnold Arboretum. The tree species observed at the Arnold Arboretum were *Acer rubrum*, *Acer saccharum*, *Aesculus flava*, *Betula alleghaniensis*, *Betula nigra*, *Carya glabra*, *Carya ovata*, *Fagus grandifolia*, *Populus deltoides*, *Quercus alba*, *Quercus rubra*, and *Tilia americana* and the shrub species were *Hamamelis virginiana*, *Vaccinium corymbosum*, and *Viburnum nudum*.

The forest site is in Petersham, MA at the Harvard Forest (42°31'53.5' N -72°11'24.1' W) and all individuals are naturally grown so provenance latitude is the same as the growing latitude. The Harvard Forest is 1446 acres and has a range of elevation of 220-410 m. We again used budburst observations—defined as BBCH stage 07 and collected by Dr. John O’Keefe weekly (O’Keefe, 2014)—across 63 individuals, all with the same provenance and growing latitude of 42.53.

The tree species observed at the Harvard Forest site were *Acer rubrum*, *Acer saccharum*, *Betula alleghaniensis*, *Fagus grandifolia*, *Fraxinus americana*, *Quercus alba* and *Quercus rubra* and the shrub species were *Acer*

pensylvanicum and *Hamamelis virginiana*. Thus, the overlapping species between the two sites include: *Acer rubrum*, *Acer saccharum*, *Betula alleghaniensis*, *Hamamelis virginiana*, *Fagus grandifolia*, *Quercus alba* and *Quercus rubra*.

We deployed 15 HOBO loggers (without radiation shields)—at approximately 1.3 m above the ground—across each site along long-term phenology observation routes (O’Keefe, 2014). We first calibrated loggers by placing them all in a growth chamber at 4°C for 24 hours and adjusted the recordings by subtracting the deviations from 4°C. We compared these HOBO logger temperatures to weather station temperatures. Weather stations at each site were on towers, at much higher elevations than 1.3 m.

HOBO loggers were deployed in October 2018 and phenology observations from the spring of 2019 were used in this study. We intended to collect phenology observations and temperature data for 2020 as well, but due to the COVID-19 pandemic we were only able to consistently collect data for 2019.

Simulations

We simulated ‘test data’ to assess inference from our models on teasing out effects of microclimate effects 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 provenance effects (i.e., there were multiple provenance latitudes at the urban arboretum site but only one at the rural forest site), (3) presence of microclimate effects (at one or both sites) accurately measured by HOBO loggers and (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 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 increased the GDD threshold for individuals at the more urban locations. To test the provenance latitude hypothesis we

made individuals from more northern provenances require fewer GDDs. To test microclimate effects, we built our climate data then added variation to this weather data to create “microclimate” effects. To test for the effect of noise, we added 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 GDD (i.e., we simulated cases where the species’ base temperature was 0°C versus 10°C) with variation in sigma (noise) around mean temperatures (i.e., 0.1°C and 1°C, where higher sigma yield higher simulated variability in daily temperatures). 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 without warming. Accuracy was evaluated as a ratio of observed GDD divided by the expected GDD, with perfect accuracy measured as 1. Values that deviate from 1 represent a percent change in inaccuracy (e.g., 1.1 is 10% inaccurate; values are never less than 1 because observed GDD must always be equal to or greater than expected GDD in order for the threshold for budburst to be met).

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. We measured GDDs by subtracting the threshold temperature (i.e., 0°C) from the mean daily temperature and then summing up these differences overtime (Asse *et al.*, 2018). If the mean temperature was below 0°C, then no GDDs were accumulated on that day. 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 17 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 \pm 50% uncertainty intervals relative to the rural, forested site using HOBO logger data from our models in the main text because these intervals are more computationally stable (Gelman *et al.*, 2014; Carpenter *et al.*, 2017). See Tables S1-S7 for 95% uncertainty intervals. In model output figures, we also report variance (i.e., the ‘sigma’ values) around major parameters from the model, which help understand partitioning of variance 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 (<https://github.com/cchambe12/microapp>). 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 found that we could accurately recover a simple effect of (1) urban sites requiring more GDDs until budburst (Figure 1a and Table S1) and (2) more northern provenance locations requiring fewer GDDs until

budburst, which was recovered in the provenance parameter (Figure 1b and Table S2).

Simulations that included microclimates at both sites recovered ‘noisier’ estimates for the method parameter with HOBO loggers requiring more GDDs until budburst. When simulating microclimate effects—thus greater variation in GDD—across the sites, we included greater variation in temperature for the HOBO logger data. Greater temperature variability led to more days at higher temperatures, so the day of budburst ultimately recorded higher GDDs, which was recovered in the negative slope of the method parameter (Figure 1c and Table S3). When we manipulated the simulations to have noisy weather station data, noise was recovered as the sigma for the method parameter (Figure 1d and Table S4) and weather stations required slightly more GDDs until budburst. When we manipulated the simulations to have noisy HOBO logger data, the output was nearly identical (Figure 1e and Table S5), but HOBO loggers required slightly more GDDs until budburst. Both the simulation for microclimate effects at both sites and the simulation for noisy (i.e., less accurate HOBO logger estimates) reported very similar results.

Simulations: GDD accuracy

We found the GDD model is less accurate with warming, and accuracy decreased at a faster rate with the lower base temperature (i.e., 0°C) than with the higher base temperature (i.e., 10°C; Figure 2). Using the 10°C base temperature, GDD accuracy was highest across all GDD thresholds and across all scenarios of warming. Without warming, the GDD model was more accurate for individuals that have high GDD thresholds and when base temperatures are higher (i.e., 10°C; Figure S1). Additionally, variability in accuracy increased with higher simulated variability in daily temperatures under warming conditions and across GDD thresholds (Figure S1 and Figure 2).

Empirical data

Mean temperature from January 1 until May 31 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 3). Using climate data from HOBO

loggers, mean spring temperature at the urban arboretum was 6.13°C and was 1.78°C at the rural forested site (Figure 3). Overall, the HOBO loggers generally recorded higher temperatures than the weather station at the urban arboretum site (with a mean difference of 1.75°C and a standard deviation of 1.03°C; Figure S2). There was greater variation in recorded temperature from the weather station at the rural forested site, though it did not typically record higher or lower temperatures than the HOBO loggers: the mean difference was 0.55°C with a standard deviation of 1.04°C (Figure S2).

Individuals at the urban arboretum site required fewer GDDs to budburst than the individuals at the rural forested site (as mentioned above, all values are given as percent and mean \pm 50% uncertainty intervals, relative to the rural forested site using HOBO logger temperature data; -9.3%, -40.75 ± 19.42 GDDs until budburst; Figure 4 and Table S6). We also found high variation in GDDs between the two methods (sigma of 17.13 GDDs until budburst) though the mean effect is close to zero (0.34%, 1.47 ± 13.69 GDDs until budburst). Weather station data at the arboretum required the fewest number of GDDs until budburst (method x site interaction: -9.94%, -43.53 ± 15.51 GDDs until budburst). This interactive effect of method x site was the strongest predictor of GDDs, even stronger than the effect of site. This is likely due to both higher temperatures and greater variation in temperatures recorded by HOBO loggers at the urban arboretum (Figure 3 and Figure S2). HOBO loggers across the two sites reported similar estimates of GDDs until budburst, whereas the weather station at the arboretum reported much lower GDDs until budburst than the rural forest weather station (Figure 5).

GDD values for species ranged from 132 to 667, with shrubs generally requiring fewer GDDs until budburst than trees (Figure S4). Our raw empirical data and model output suggests shrubs require fewer GDDs (i.e., mean of 386 GDD) until budburst than trees (mean of 407 GDD; Figure S4). At the rural site, species and functional-type (tree versus shrub) GDD estimates were consistent across the climate data method used, whereas there was a bigger difference between the two methods at the urban arboretum. Individuals across all species at the rural forest site required more GDDs until budburst than at the urban arboretum (Figure S5a and b), but there was large variation in species requirements across the two climate data methods, especially

for the raw data (Figure S5c). The model output estimates comparing the two climate data methods show very little difference in GDD requirements for all species though there is large variation around the estimates (Figure S5d).

Finally, we found no major effect of provenance latitude on GDDs until budburst though there was a slightly positive trend with higher provenance latitudes requiring more GDDs until budburst (16.17 ± 13.64 GDDs until budburst; Figure S3 and Table S7), but the variance around species was large (sigma of 14.45 GDDs until budburst). The effect of method on GDDs until budburst was close to zero (-5.7 ± 8.52 GDDs until budburst). The interaction of provenance by method was also close to zero (-2.44 ± 13.25 GDDs until budburst), but the variance across species was large (sigma of 12.89 GDDs until budburst).

Discussion

Our study assessed the effects of an urban arboretum versus a more rural forested site coupled with the effect of climate data type (i.e., weather station versus HOBO logger temperature data) on estimated GDD until budburst. We found the urban site was warmer, but this did not translate to individuals requiring more GDDs as we hypothesized, but rather fewer GDDs until budburst. Our results additionally suggest there was a strong microclimate effect (apparent by the large variation in GDD with method). Though these effects varied by site: HOBO loggers at the urban arboretum generally recorded higher temperatures than the weather station, and HOBO loggers at the rural forested site recorded more variation in temperatures than the associated weather station. Regardless of site or climate data method used, however, trees consistently required more GDDs until budburst than shrubs, which has important implications for model forecasts. The effect of provenance latitude did not follow any clear pattern. This could be due to the weak effect of latitude on spring phenology (Gauzere *et al.*, 2017), or our limited sample, especially in its range of provenance latitudes (Figure S6). Given the potential for latitude to have a small effect size, we suggest future studies interested in teasing out provenance effects should include a greater range of latitudes and/or more sampled trees across this range, compared to our study.

Variation across and within sites and among species suggests important variation for forecasting

Our finding that individuals growing in urban environments require fewer GDDs—which was consistent across species—contributes to increasing evidence that trees in urban areas may respond differently than those in forested, rural areas. This finding of urban sites requiring fewer GDDs is broadly in line with one recent study that found urban sites have a lower temperature sensitivity compared to colder rural sites (Meng *et al.*, 2020). This means that long-term records and experiments conducted in urban areas may not be transferable to larger scales, including in models that incorporate forested rural areas.

The lower GDD requirement in the urban arboretum could be due to higher over-winter chilling. While numerous studies assume warming will decrease chilling (Luedeling *et al.*, 2011; Fu *et al.*, 2015; Asse *et al.*, 2018), actual effects of warming depend on the range of temperatures over which plants accumulate chilling. 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 can be equally important as over-winter temperatures. Additionally, if 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 suggest 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.

Our finding of lower GDDs until budburst at the urban site, however, depends on the method of recording climate data. We found 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 consistently lessen the urban effect—as we saw here. Our results suggest that microclimatic effects, and the location of the weather station, may have major impacts on our interpretation of how different GDD requirements are for trees in urban versus rural sites.

Our results additionally indicate there were microclimatic effects at both sites, but a larger effect of micro-

climate at the urban arboretum site. At the rural forested site, there was greater variation in temperatures recorded from the HOBO loggers than the weather station, but overall the climate data from the HOBO logger and the weather station were more similar. Further, the difference in temperatures between the two methods at the arboretum occurred at biologically meaningful temperatures: the weather station tended to record cooler temperatures than most of the HOBO loggers (Figure 3), putting the weather station temperatures often close to or under 0°C at the same time that some HOBO loggers were above 0°C, the threshold for accumulating GDD in many forest tree models (Man & Lu, 2010). This effect may be due to canopy differences between the two sites, with the urban arboretum having a generally open-canopy and high variation of species (and thus canopy types) across space, versus a typically closed-canopy, rural forest where species composition was more consistent; or due to effects of roads and other urban structures at the urban arboretum (Stabler *et al.*, 2005; Erell *et al.*, 2012; Dimoudi *et al.*, 2013). Additionally, at the rural site, the weather station is situated in the middle of the forest, whereas the weather station in the arboretum is located on a hill towards the edge of the site. Whatever the cause, our results suggest that the collection method for weather data impacts GDD models.

Accuracy of GDD models varies predictably with species and daily temperature

Our results support concerns that GDD models may not be appropriate for the future with warming (Man & Lu, 2010). With warming, GDDs accumulate at a higher daily rate, which will reduce accuracy of determining the actual threshold for budburst phenology. Additionally, accuracy is greater for later-active species because they have higher GDD thresholds than early-active species. Our simulations show that a higher GDD threshold means a lower accuracy (GDD observed to GDD expected ratio) because being off by a day is a small effect for higher GDD threshold species (and hence greater days to budburst) than for lower GDD threshold species. We also found through simulations that species with a higher base temperature will be estimated more accurately through GDD models (given the same GDD threshold).

These results—across warming, GDD thresholds, and base temperatures—all highlight an intrinsic reality to

GDD models: because they are accumulated each day, they are more accurate when there are more days to an event. More days to an event can occur via lower daily temperatures, higher GDD thresholds or higher base temperatures. Accuracy also depends on climate variability because high variability means some days will accumulate GDDs quickly, which can override the trend we see of higher GDD threshold being more accurate. In reality temperature variance likely changes over the spring (Qu *et al.*, 2014), rendering climate change effects even harder to decipher. These issues are not inherently unique to GDD models—any biological process dependent on temperature that is measured over days becomes less accurate with warming—but they highlight realities to using GDD models for forecasts, including how accuracy may inherently be lower for warmer areas, and certain species.

Accurately attributing observed variation requires greater insights into climate methods and phenology

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. Thus, determining which methods are most accurate is the first step to establishing fine-scale climatic variation and better forecasts of phenology with climate change. Our results show that large fluctuations in spring temperatures leads to higher GDDs until budburst since GDDs accumulate faster with higher spring temperature variability—and, ultimately, more frequent high temperature days. Our simulations suggest teasing out noise versus microclimate effects can be difficult, but when considering climate data and GDD until budburst estimates together, our empirical results suggest there are microclimatic effects that vary by site. Overall our work suggests we must better understand what underlies temperature variability (i.e., inaccurate methods or microclimate effects) to improve our phenology forecasts based on GDD.

Our results suggest that more accurately modeling GDDs under climate change will require additional studies of how local climate determines phenology. More research on climate methods, including specific studies of HOBO logger location in the canopy and to apply these treatments next to both weather stations and to

the trees or shrubs of interest, may be especially useful. We also suggest further studies on the effects of radiation shields on overall precision and accuracy of the temperatures recorded (da Cunha, 2015). Understanding how important radiation shields are for GDD models, however, requires a better understanding of what temperatures are most important to plant phenology (e.g., bud temperature, including influences of bud color and structure and their interaction with solar radiation, versus air temperature). Additionally, as many ecosystem models predict phenology by functional type rather than species, more studies that discern differences in GDD requirements across functional groups are crucial. Our results suggest we may fundamentally estimate early-active species, such as shrubs, less accurately with GDD models, and highlight the need to incorporate this uncertainty. Finally, as others have (Duputié *et al.*, 2015; Chuine *et al.*, 2016), we propose the continued improvement of mechanistic models of spring phenology could help move away from GDD models. Continuing evidence suggests that simple GDD models may not be as accurate as models that incorporate other phenological cues beyond spring warming, such as chilling and photoperiod (Keenan *et al.*, 2020). Incorporating these cues usefully in forecasts may be difficult, as our results highlight the complexity of even apparently simple GDD models when estimating across space and species.

Acknowledgments

We would like to thank all of the Arnold Arboretum Tree Spotters and grounds crew for observing and maintaining the trees, with a special thanks to S. Mrozak, D. Schissler, P. Thompson and K. Stonefoot for their continued dedication to the Tree Spotters program. We dedicate a special thank you to Dr J. O’Keefe for his work and observations at the Harvard Forest. We also want to thank D. Buonaiuto, W. Daly, 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

- Abendroth, L.J., Miguez, F.E., Castellano, M.J. & Hatfield, J.L. (2019) Climate warming trends in the u.s. midwest using four thermal models. *Agronomy Journal* **111**, 3230–3243.
- Aitken, S.N. & Bemmels, J.B. (2015) Time to get moving: assisted gene flow of forest trees. *Evolutionary Applications* **9**, 271–290.
- Asse, D., Chuine, I., Vitasse, Y., Yoccoz, N.G., Delpierre, N., Badeau, V., Delestrade, A. & Randin, C.F. (2018) Warmer winters reduce the advance of tree spring phenology induced by warmer springs in the alps. *Agricultural and Forest Meteorology* **252**, 220–230.
- 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.

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

391 Bolton, D.K. & Friedl, M.A. (2013) Forecasting crop yield using remotely sensed vegetation indices and crop
392 phenology metrics. *Agricultural and Forest Meteorology* **173**, 74–84.

393 Bonhomme, R. (2000) Bases and limits to using ‘degree.day’ units. *European Journal of Agronomy* **13**, 1–10.

394 Carpenter, B., Gelman, A., Hoffman, M., Lee, D., Goodrich, B., Betancourt, M., Brubaker, M.A., Guo, J.,
395 Li, P. & Allen, R. (2017) Stan: A probabilistic programming language. *Journal of Statistical Software* **76**,
396 10.18637/jss.v076.i01.

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

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

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

403 Chuine, I., Bonhomme, M., Legave, J.M., García de Cortázar-Atauri, I., Charrier, G., Lacointe, A. & Améglio,
404 T. (2016) Can phenological models predict tree phenology accurately in the future? the unrevealed hurdle
405 of endodormancy break. *Global Change Biology* **22**, 3444–3460.

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

409 Crimmins, M.A. & Crimmins, T.M. (2019) Does an early spring indicate an early summer? relationships
410 between intraseasonal growing degree day thresholds. *Journal of Geophysical Research: Biogeosciences* **124**,
411 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.
- da Cunha, A.R. (2015) Evaluation of measurement errors of temperature and relative humidity from hobo data logger under different conditions of exposure to solar radiation. *Environmental Monitoring and Assessment* **187**, 236.
- de Rességuier, L., Mary, S., Le Roux, R., Petitjean, T., Quénot, 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.
- Dimoudi, A., Kantzioura, A., Zoras, S., Pallas, C. & Kosmopoulos, P. (2013) Investigation of urban microclimate parameters in an urban center. *Energy and Buildings* **64**, 1–9.
- Dosmann, M.S. (2006) Research in the garden: Averting the collections crisis. *The Botanical Review* **72**, 207–234.
- Duputié, A., Rutschmann, A., Ronce, O. & Chuine, I. (2015) Phenological plasticity will not help all species adapt to climate change. *Global Change Biology* **21**, 3062–3073.
- Erell, E., Pearlmutter, D. & Williamson, T. (2012) *Urban microclimate: designing the spaces between buildings*. Routledge.
- 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.
- Finn, G., Straszewski, A. & Peterson, V. (2007) A general growth stage key for describing trees and woody plants. *Annals of Applied Biology* **151**, 127–131.
- Fu, Y.H., Zhao, H., Piao, S., Peaucelle, M., Peng, S., Zhou, G., Ciais, P., Huang, M., Menzel, A., Peñuelas, J. & et al. (2015) Declining global warming effects on the phenology of spring leaf unfolding. *Nature* **526**, 104–107.

- 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.
- Klosterman, S., Hufkens, K. & Richardson, A.D. (2018) Later springs green-up faster: the relation between onset and completion of green-up in deciduous forests of North America. *International Journal of Biometeorology* **62**, 1645–1655.
- Kumudini, S., Andrade, F.H., Boote, K.J., Brown, G.A., Dzotsi, K.A., Edmeades, G.O., Gocken, T., Goodwin, M., Halter, A.L., Hammer, G.L., Hatfield, J.L., Jones, J.W., Kemanian, A.R., Kim, S.H., Kiniry, J., Lizaso, J.I., Nendel, C., Nielsen, R.L., Parent, B., Stöckle, C.O., Tardieu, F., Thomison, P.R., Timlin, D.J., Vyn, T.J., Wallach, D., Yang, H.S. & Tollenaar, M. (2014) Predicting maize phenology: Intercomparison of functions for developmental response to temperature. *Agronomy Journal* **106**, 2087–2097.
- 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.
- Luedeling, E., Girvetz, E.H., Semenov, M.A. & Brown, P.H. (2011) Climate change affects winter chill for temperate fruit and nut trees. *PLOS ONE* **6**, 1–13.
- 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., Geraldès, 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 in *populus 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.
- Montgomery, R.A., Rice, K.E., Stefanski, A., Rich, R.L. & Reich, P.B. (2020) Phenological responses of temperate and boreal trees to warming depend on ambient spring temperatures, leaf habit, and geographic range. *Proceedings of the National Academy of Sciences* **117**, 10397–10405.
- 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.
- O’Keefe, J. (2014) Phenology of woody species at Harvard Forest since 1990.

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

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

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

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

493 Qu, M., Wan, J. & Hao, X. (2014) Analysis of diurnal air temperature range change in the continental united
494 states. *Weather and Climate Extremes* **4**, 86–95.

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

497 Richardson, A.D., Keenan, T.F., Migliavacca, M., Ryu, Y., Sonnentag, O. & Toomey, M. (2013) Climate
498 change, phenology, and phenological control of vegetation feedbacks to the climate system. *Agricultural*
499 *and Forest Meteorology* **169**, 156 – 173.

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

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

505 Schwartz, M.D., Hanes, J.M. & Liang, L. (2013) Comparing carbon flux and high-resolution spring pheno-
506 logical 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.
- Stabler, L.B., Martin, C.A. & Brazel, A.J. (2005) Microclimates in a desert city were related to land use and vegetation index. *Urban Forestry Urban Greening* **3**, 137–147.
- 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.
- Vico, G., Karacic, A., Adler, A., Richards, T. & Weih, M. (2021) Consistent poplar clone ranking based on leaf phenology and temperature along a latitudinal and climatic gradient in northern europe. *BioEnergy Research* **14**, 445–459.
- 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.

- 531 Zhang, R., Lin, J., Wang, F., Shen, S., Wang, X., Rao, Y., Wu, J. & Hänninen, H. (2021) The chilling require-
532 ment of subtropical trees is fulfilled by high temperatures: A generalized hypothesis for tree endodormancy
533 release and a method for testing it. *Agricultural and Forest Meteorology* **298-299**, 108296.
- 534 Ziter, C. & Turner, M.G. (2018) Current and historical land use influence soil-based ecosystem services in an
535 urban landscape. *Ecological Applications* **28**, 643–654.
- 536 Zohner, C.M., Benito, B.M., Svenning, J.C. & Renner, S.S. (2016) Day length unlikely to constrain climate-
537 driven shifts in leaf-out times of northern woody plants. *Nature Climate Change* **6**, 1120–1123.
- 538 Zohner, C.M. & Renner, S.S. (2014) Common garden comparison of the leaf-out phenology of woody species
539 from different native climates, combined with herbarium records, forecasts long-term change. *Ecology Let-*
540 *ters* **17**, 1016–1025.

Tables and Figures

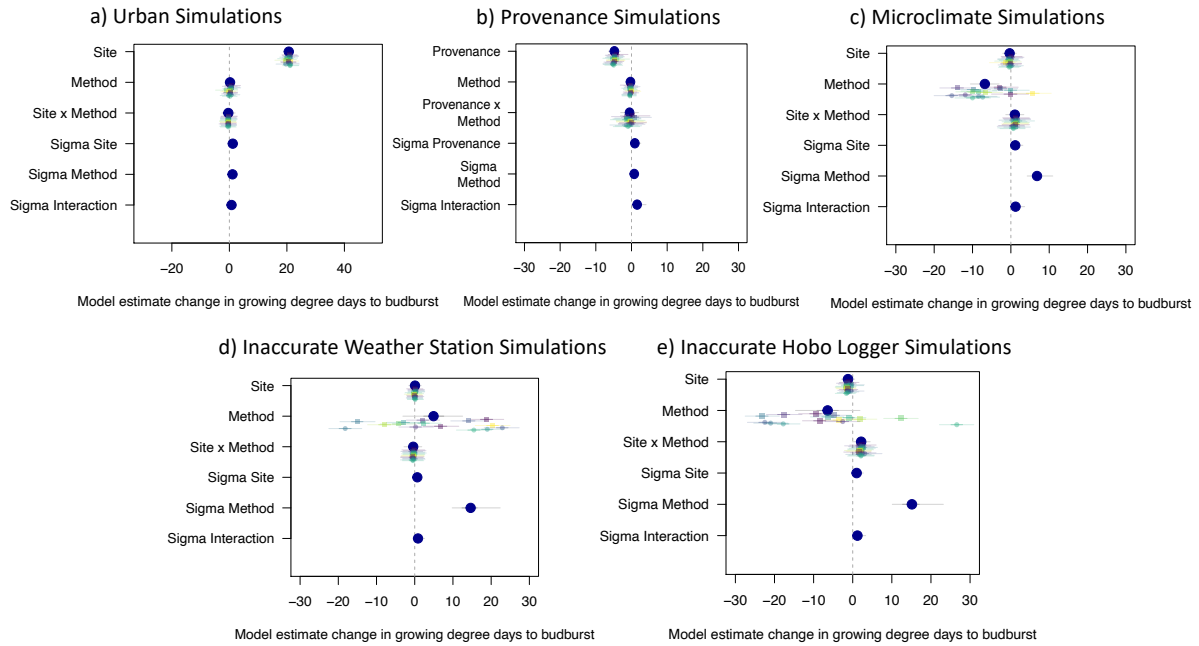


Figure 1: Simulations: we show (a) urban sites requiring more GDDs, (b) more northern provenance latitudes requiring fewer GDDs, (c) microclimate effects, (d) less accurate weather station data and (e) less accurate HOBO logger data, which looks similar to (c). We show the effects of site (urban versus rural) and method (weather station versus HOBO loggers) in (a), (b), (d) and (e). The intercept represents the HOBO logger data for the rural forested site. More positive values indicate more GDDs required for budburst whereas more negative values suggest fewer GDDs required. Dots and thin lines show means and 90% uncertainty intervals and thick lines show 50% uncertainty intervals. See Tables S1, S3, S2, S4 and S5 for full model output.

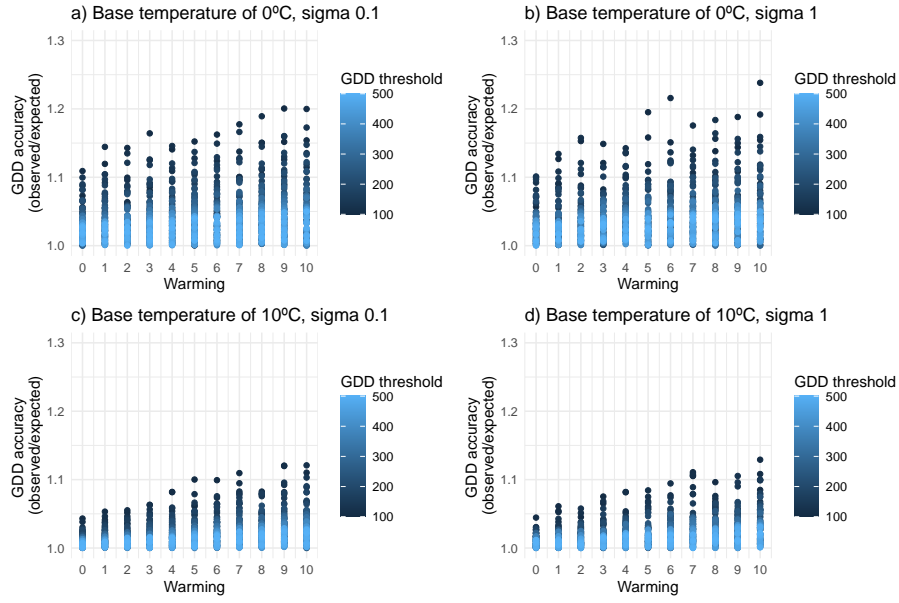


Figure 2: Using simulated data, we show how GDD measurement accuracy changes with warming (i.e., from 0°C to 10°C) using a base temperature of (a) 0°C and a sigma of 0.1°C, (b) 0°C and a sigma of 1°C, (c) 10°C and a sigma of 0.1°C and (d) 10°C and a sigma of 1°C. GDD accuracy is measured as the observed GDD divided by the expected GDD. Values closest to 1 are most accurate, with values deviating from 1 representing a percent change in inaccuracy (e.g., 1.1 is 10% inaccurate). Observed GDD must always be equal to or greater than expected GDD in order for the threshold for budburst to be met.

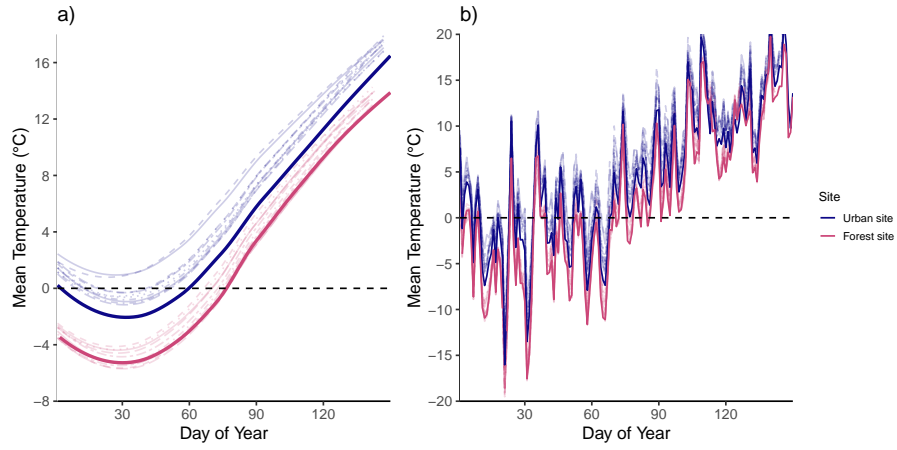


Figure 3: Here we show a breakdown of the climate data across the two sites with darker lines representing weather station data and the lighter, more transparent lines of varying line types representing the HOBO loggers: a) a series of smoothing splines of mean temperature with 90% uncertainty interval and b) actual mean temperature.

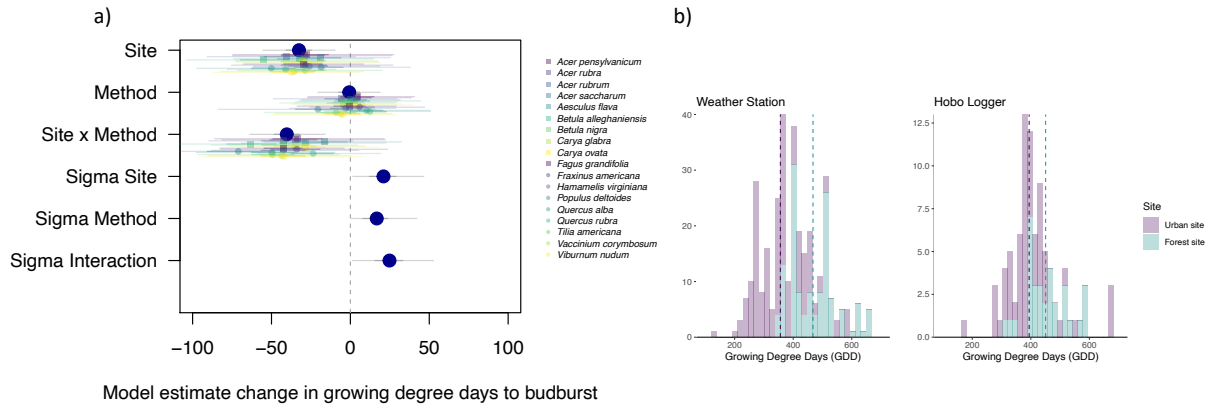


Figure 4: Empirical Data: we show (a) the main effects and variance (sigma) of site (urban versus rural) and climate data method (weather station versus HOBO loggers), as well as their interaction, on GDDs until budburst. The intercept represents the HOBO logger data for the rural forested site. More 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.

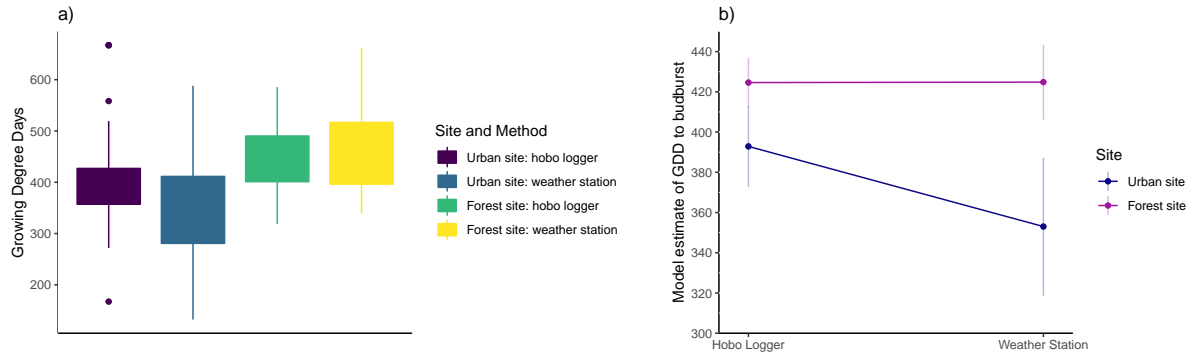


Figure 5: We show estimated effects, from a Bayesian hierarchical model, of site (urban arboretum site versus forested rural site) by climate data method (weather station data versus HOBO logger data) on GDDs until budburst (a) as a boxplot across each method and site combination using raw data and (b) using model output to show the mean estimates for each site and method with 50% uncertainty intervals shown as error bars.