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:

[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]The Nature Conservancy, Durham, NC 27701 USA;

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

*Corresponding author: 248.953.0189; c.chamberlain@tnc.org; 0000-0001-5495-3219

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

Abstract

Predicting spring phenology in temperate forests is critical for forecasting important processes such as carbon storage. 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 is constant across diverse landscapes, but increasing evidence suggests otherwise. Shifts in climate with anthropogenic warming, may alter the required GDD. Variation in climate across space may also lead to variation in GDD requirements, with recent studies suggesting that fine-scale spatial variation in climate may matter to phenology. Here, we combine simulations, observations from an urban and a rural site, and Bayesian hierarchical models to assess how consistent GDD models of budburst are across species and space.

We built GDD models using two different methods to measure climate data: on-site weather stations and local dataloggers. We find that estimated GDD thresholds can vary up to 20% across sites and methods. Our results suggest our studied urban site requires fewer GDDs until budburst and may have stronger microclimate effects than the studied rural site, though these effects depend on the method to measure climate. Further, we find that GDD models are less accurate for early-active species and may become less accurate with warming. Our results suggest that local-scale forecasts based on GDD models for spring phenology should incorporate these inherent accuracy issues of GDD models, alongside the variations we found across space, species and warming. Testing whether these issues persist at larger spatial scales could improve forecasts for temperate forests.

1 Introduction

Understanding and predicting spring plant phenology in temperate deciduous forests is critical as it both shapes community structure and influences resource and forest management (1; 2). Climate change and urbanization are advancing spring timing—such as budburst and leafout, which are strongly cued by temperature, resulting in longer growing seasons (3). These shifts in growing seasons ultimately impact ecosystem services and forest management.

Spring budburst timing in particular can have cascading effects on pollinators (4; 5), albedo (6), and carbon dynamics (7). 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 (8). Because of this, forecasting phenology accurately with climate change is a major and important aim across several fields of science including agronomy, ecology, evolution and hydrology (9; 10; 11; 12).

One major forecasting method widely accepted and utilized across these fields is the growing degree day (GDD) model, which allows researchers to track thermal sums (or heat accumulation) to predict spring budburst (13; 14; 15; 16; 17). While researchers have refined thermal sum models for some well studied species, such as herbaceous crops, to be more precise (18; 19), the GDD model remains widely accepted, and has been far more widely adapted for wild tree and shrub species (11; 20; 21). An often-used form of the model takes the mean daily temperature, subtracts this value from the base temperature—often 0°C for forest trees (as estimates are proven to be more accurate, 22)—sums these temperatures each day and accumulates them until leafout. Different species generally require a different number of GDDs to leaf out, with early-leafout species requiring less than later-leafout species (23; 17; 24). 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 the same event (25). Understanding the complexity of the apparently simple GDD model 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 across individual plants and locations, but increasing evidence suggests it may not be. Due to plasticity the same individual could leaf-out at different GDD accumulations. This could be driven by additional cues underlying leafout, as decades of work show that chilling—related to winter temperatures—and photoperiod can shift the GDD a plant needs to trigger a phenological event (26; 27; 28).

Climate helps determine the role of chilling and photoperiod—and, thus the required GDD (25; 29). On a large scale climate gradients across space (i.e., latitudinal or continentality effects) and gradients due to anthropogenic impact may thus alter estimated GDD. Urbanization has led to the formation of urban heat islands, which can affect plant phenology and lead to earlier spring leafout—and varying GDD thresholds—due to smaller chilling effects through warmer winters (30).

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 (31; 32). Additionally urban sites often house arboreta or botanical gardens that typically contribute long-term phenology records (33) or are used for experiments on phenology (34). Recent work suggests arboreta and botanical gardens offer a unique lens to investigate climate change and local adaptation because these sites often incorporate varying seed sources—or provenance (i.e., origin) locations—thus mimicking common garden experiments (35). 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.

Provenance effects on spring phenology are related to phenology's genetic component. Though local adaptation in spring phenology is generally much lower than in fall phenology (36; 37; 38), the required chilling, photoperiod and GDD for spring events can vary by population (39; 40; 29). 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 (28), 2) photoperiod will slow or constrain range expansion (41), 3) 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 (43). Additionally, there is evidence that 5) artificial light can counteract the photoperiod limitation (44), which could contribute to urban level effects in combination with latitudinal effects. Many arboreta keep diligent acquisition records, providing visitors and scientists information on seed sources (45), and the potential to test such provenance effects.

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 site or up to 6.6°C within 1 km spatial units in northern Europe, 46; 47). Increasing evidence suggests that fine-scale climate may matter to phenology (48; 49). To facilitate scaling and minimize error due to these fine-scale climatic effects, which we refer to as microclimate effects, researchers often deploy standalone small, local dataloggers—such as HOBO sensors—which may provide higher resolution weather data (50; 51).

Here, we aimed to address the following hypotheses using a case study approach where we compared trees in an urban arboretum (which has multiple provenances represented) to a rural forested site (which only has one provenance represented): 1) required GDD in an urban arboretum will vary from a rural forested site, where we predicted that more GDDs would be required to trigger leafout at the urban site due to lower chill accumulation, 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. To better interpret our results and GDD models we used simulations, which have been used previously to understand the value of GDD models to predicting budburst (e.g., 52).

2 Material and methods

2.1 Sites and species

We chose two sites—one urban arboretum and one rural forest—with overlapping species and climates to compare the number of GDDs 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 numerous woody plant taxa from North America, Europe and Asia and has an elevation range 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, see 53) 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 (54).

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 rural 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 (55)—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 local dataloggers (HOBO from Onset Corporation) at both the Arnold Arboretum and the Harvard Forest site along long-term phenology observation routes (55; 54). Because we were interested in plant-experienced climate we placed loggers at approximately 1.3 m above the ground without radiation shields. 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.

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

2.2 Simulations

We simulated 'test data' (sometimes referred to as 'artificial data,' see 52) 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) more GDDs would be required to trigger leafout at urban environments, 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. Our models of microclimate effects assume station and local (HOBO) dataloggers collect equally accurate data, but this may not be true either because of the sensors, or because the plants experience important differences in weather compared to the loggers. We were interested in whether such differences in accuracy would manifest in models differently than microclimate effects. Thus we also tested: 4) weather stations or HOBO loggers are effectively 'noisier' (less accurate) data for GDD models compared to the other.

To run our simulations, assumed each species required a different GDD threshold (drawing each species' requirement from a normal distribution). We then simulated climate data by again establishing a distribution around a mean temperature for each site, with mean temperatures matching real climate data. Using this climate data, we found the day of budburst when the unique GDD threshold was met for each individual. To test that plants located in urban locations require the accumulation of more GDDs, 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 temperatures 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 (i.e., the degree of 'noise' or variation) 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 (56), version 2.19.2, in R (57), version 3.3.1, we estimated the effects of urban, provenance, and method and all two-way interactions as predictors on GDDs until budburst. We measured GDDs for both the empirical models and the simulation models by subtracting the base temperature (i.e., 0°C) from the mean daily temperature and then summing up these differences over days (58) since January 1. If the mean temperature was below 0°C, then no GDDs were accumulated on that day. We additionally tested accuracy across varying baseline temperatures in our simulations. 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 (59). We report means ± 50% uncertainty intervals in the main text because these intervals are more computationally stable (59; 60). See Tables S1-S7 for 95% uncertainty intervals. All estimates (unless otherwise noted) are relative to the rural, forested site using local datalogger (HOBO) data from our models. In model output figures, we also report variance (i.e., the 'sigma' values) around major parameters from the model, which describe partitioning of variance within the model (59). We standardized all model predictors to allow direct comparison across model estimates and all estimates (unless otherwise noted) are relative to the rural, forested site using local datalogger (HOBO) data from our models.

Shiny application

To show the above simulations, real data and forecasts in one location we use a Shiny Application (`https://impactforecasting.shinyapps.io/ecomodels_GDD/`). Using the R package 'shiny' (61), 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.

3 Results

3.1 Simulations

We found that we could accurately recover two simple effects of 1) urban sites requiring more GDDs until budburst (Figure 1**a** and Table S1) and 2) more northern provenances requiring fewer GDDs until budburst (Figure 1**b** and Table S2).

Simulations for microclimate effects (at both sites) and the simulations for noisy (i.e., less accurate HOBO logger estimates) reported similar results. Including microclimates at both sites led to more variable estimates for the method parameter with local (HOBO) dataloggers requiring more GDDs until budburst. This occurred because when we simulated microclimate effects across the sites we included greater variation in temperature for the HOBO logger data, which led to more days at higher temperatures, and ultimately a day of budburst that recorded higher GDDs (this was recovered in the negative slope of the method parameter, see Figure 1**c** 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 1**d** and Table S4) and weather stations required slightly more

GDDs until budburst. When we manipulated the simulations to have noisy local (HOBO) dataloggers, the output was nearly identical (Figure 1**e** and Table S5), but HOBO loggers required slightly more GDDs until budburst.

3.2 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).

3.3 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 local (HOBO) dataloggers 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 local (HOBO) dataloggers: the mean difference was 0.55°C with a standard deviation of 1.04°C (Figure S2).

Individual plants at the urban arboretum site required fewer GDDs to budburst than individuals at the rural forested site (as mentioned above, all values are given as percent and mean ± 50% uncertainty intervals, relative to the rural forested site using local (HOBO) datalogger 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 $\pm$ 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 $\pm$ 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 local (HOBO) dataloggers at the urban arboretum (Figure 3 and Figure S2). Local (HOBO) dataloggers 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 thresholds 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 6**a** and **b**), but there was large variation in species requirements across the two climate data methods, especially for the raw data (Figure 6**c**). 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 6**d**).

Finally, we found no major effect of provenance latitude on GDDs until budburst though there was a slightly positive trend, with more northern provenance latitudes requiring more GDDs until budburst (16.17 $\pm$ 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 $\pm$ 8.52 GDDs until budburst). The interaction of provenance by method was also close to zero (-2.44 $\pm$ 13.25 GDDs until budburst), but the variance across species was large (sigma of 12.89 GDDs until budburst).

4 Discussion

Our case study approach, which compared GDD requirements necessary to trigger leafout between an urban arboretum and a rural forested site, and simulations suggest important variation across locations, climate data and species when using GDD to model budburst in forest trees. Similar to other research (30), we found the urban site was warmer, but this did not translate to individuals requiring more GDDs as we hypothesized, but rather fewer GDDs. Our simulations showed that teasing out microclimate effects from noisy weather data can be difficult. However, using these simulations to help interpret our empirical results, our findings suggest microclimatic effects at both locations, with a larger effect of microclimate at the urban arboretum site. Regardless of site or climate data method used, however, shrubs consistently required fewer GDDs until budburst than trees, which has important implications for model forecasts. Our simulations highlight, however, that early-active species, such as shrubs will inherently be estimated less accurately than later-leafout species using GDD models (62).

These findings have relevance to local-scale forecasting. Land managers of temperature forests working at local to smaller regional scales may want to consider how microclimatic, species-level effects and GDD accuracy issues could impact their planning with climate change. As our climatic and species-level data were gathered at a relatively small spatial scale, extending these findings to larger spatial scales (e.g., with remote sensing etc.) to test whether simple approaches like GDD thresholds may still perform reasonably well is an important next step.

While there is growing interest to use arboreta to understand provenance effect (35), we did not find any clear pattern with provenance latitude. This could be due to the weak effect of latitude on spring phenology (43), or our limited sample, especially in its range of provenance latitudes (Figure S5). 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.

4.1 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 (30). This suggests 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 arboretum could be due to shifts in other phenological cues in urban settings. While light is highly altered in urban settings (63), increasing evidence suggests it is a weak cue for spring tree phenology compared to chilling (64). Higher over-winter chilling in the urban site, however, could explain our findings. While numerous studies assume warming will decrease chilling (65; 66; 58), 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 (67)—or even up to 15°C in subtropical trees (68)—but the duration of winter can be equally important. Additionally, if temperatures are below the chilling accumulation threshold—which may occur at cooler sites where temperatures are more frequently too cold to accumulate chilling—then we would 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 in terms of chilling accumulation.

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 was weaker when estimated using local (HOBO) dataloggers at both sites. Further studies that investigate more rural and associated urban sites are necessary to test if local dataloggers 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. Additionally, different types of temperature logging equipment could vary in accuracy. Future studies should assess multiple types of equipment at each site to better estimate variation.

Our results additionally indicate microclimatic effects at both sites, but a larger effect of microclimate at the urban arboretum site. At the rural forested site, there was greater variation in temperatures recorded from the local (HOBO) dataloggers than the weather station, but overall the climate data from the local dataloggers 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 local dataloggers (Figure 3), putting the weather station temperatures often close to or under 0°C at the same time that some local dataloggers were above 0°C, the threshold for accumulating GDD in many forest tree models (22), and the threshold we used here. 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 our studied closed-canopy, rural forest where species composition was more consistent; or due to effects of roads and other urban structures at the urban arboretum (69; 70; 71). 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. As we have only one of each location type and a single year of data, our results should be interpreted cautiously; however, they suggest that the collection method for weather data impacts phenology models, which is echoed by other work comparing weather collection methods (e.g., 49).

4.2 Accuracy of GDD models varies predictably with species and daily temperature

Our results support concerns that predicting from GDD models may not be appropriate given continued warming (22). With warming, GDDs accumulate at a higher daily rate, which will reduce accuracy of determining the actual threshold for budburst phenology (25). Additionally, accuracy is greater for later-active species because they have higher GDD thresholds than early-active species. Our simulations show that a lower GDD threshold means a lower accuracy (GDD observed:GDD expected) 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 thresholds being more accurate. In reality temperature variance likely changes over the spring (72), 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 early-active species.

4.3 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 (49). 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, and that we must better understand what underlies temperature variability (i.e., inaccurate methods or microclimate effects) to improve our phenology forecasts based on GDD.

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 that compare results using local dataloggers at different locations 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 (73). 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, 74). 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.

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. We would also like to thank the anonymous reviewers for their thoughtful feedback and contributions to making this manuscript more accessible.

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

[1] Morellato, L. P. C., Alberton, B., Alvarado, S. T., Borges, B., Buisson, E., Camargo, M. G. G., Cancian, L. F., Carstensen, D. W., Escobar, D. F., Leite, P. T., Mendoza, I., Rocha, N. M., Soares, N. C., Silva, T. S. F., Staggemeier, V. G., Streher, A. S., Vargas, B. C., and Peres, C. A. *Biological Conservation* **195**, 60–72 (2016). https://doi.org/10.1016/j.biocon.2015.12.033.

[2] Silvestro, R., Rossi, S., Zhang, S., Froment, I., Huang, J. G., and Saracino, A. *Forest Ecology and Management* **436**, 21–26 (2019). https://doi.org/10.1016/j.foreco.2019.01.005.

[3] Chuine, I., Aitken, S. N., and Ying, C. C. *Canadian Journal of Forest Research* **31**(8), 1444–1455 Aug (2001). 10.1139/x01-072.

[4] Boggs, C. L. and Inouye, D. W. *Ecology Letters* **15**(5), 502–508 Mar (2012). 10.1111/j.1461-0248.2012.01766.x.

[5] Pardee, G. L., Inouye, D. W., and Irwin, R. E. *Global Change Biology* **24**(2), 848–857 Sep (2017). 10.1111/gcb.13865.

[6] Williamson, S. N., Barrio, I. C., Hik, D. S., and Gamon, J. A. *Global Change Biology* **22**(11), 3621–3631 (2016). 10.1111/gcb.13297.

[7] Richardson, A. D., Keenan, T. F., Migliavacca, M., Ryu, Y., Sonnentag, O., and Toomey, M. *Agricultural and Forest Meteorology* **169**, 156 – 173 (2013). https://doi.org/10.1016/j.agrformet.2012.09.012.

[8] 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., and et al. *Nature Climate Change* **4**(7), 598–604 Jun (2014). 10.1038/nclimate2253.

[9] Moorcroft, P. R., Hurtt, G. C., and Pacala, S. W. *Ecological Monographs* **71**(4), 557–585 (2001).

[10] Bolton, D. K. and Friedl, M. A. *Agricultural and Forest Meteorology* **173**, 74–84 (2013). https://doi.org/10.1016/j.agrformet.2013.01.007.

[11] Yu, M., Wang, G., and Chen, H. *Journal of Advances in Modeling Earth Systems* **8**(1), 370–386 (2016).

[12] Taylor, S. D. and White, E. P. *Ecological Applications* **30**(1), e02025 (2020). https://doi.org/10.1002/eap.2025.

[13] Schwartz, M. D., Ahas, R., and Aasa, A. *Global Change Biology* **12**(2), 343–351 (2006). 10.1111/j.1365-2486.2005.01097.x.

[14] Vitasse, Y., Francois, C., Delpierre, N., Dufrene, E., Kremer, A., Chuine, I., and Delzon, S. *Agricultural and Forest Meteorology* **151**, 969–980 (2011).

[15] 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., and et al. *Ecosystems* **15**(8), 1283–1294 Sep (2012). 10.1007/s10021-012-9584-5.

[16] Phillimore, A. B., Proios, K., O'Mahony, N., Bernard, R., Lord, A. M., Atkinson, S., and Smithers, R. J. *Journal of Ecology* **101**(3), 774–783 Apr (2013). 10.1111/1365-2745.12067.

[17] Crimmins, M. A. and Crimmins, T. M. *Journal of Geophysical Research: Biogeosciences* **124**(8), 2628–2641 (2019). 10.1029/2019JG005297.

[18] 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., and Tollenaar, M. *Agronomy Journal* **106**(6), 2087–2097 2022/09/18 (2014). https://doi.org/10.2134/agronj14.0200.

[19] Abendroth, L. J., Miguez, F. E., Castellano, M. J., and Hatfield, J. L. *Agronomy Journal* **111**(6), 3230–3243 2022/09/18 (2019). https://doi.org/10.2134/agronj2019.02.0118.

[20] Klosterman, S., Hufkens, K., and Richardson, A. D. *International Journal of Biometeorology* **62**, 1645–1655 May (2018). 10.1007/s00484-018-1564-9.

[21] Montgomery, R. A., Rice, K. E., Stefanski, A., Rich, R. L., and Reich, P. B. *Proceedings of the National Academy of Sciences* **117**(19), 10397–10405 Apr (2020). 10.1073/pnas.1917508117.

[22] Man, R. and Lu, P. *Canadian Journal of Forest Research* **40**(9), 1815–1820 Sep (2010). 10.1139/x10-129.

[23] Fahey, R. T. *Ecosphere* **7**(2), e01209 (2016). 10.1002/ecs2.1209.

[24] Marquis, B., Bergeron, Y., Simard, M., and Tremblay, F. *Frontiers in Plant Science* **11** Jul (2020). 10.3389/fpls.2020.01031.

[25] Bonhomme, R. *European Journal of Agronomy* **13**(1), 1–10 (2000). https://doi.org/10.1016/S1161-0301(00)00058-7.

[26] Basler, D. and Körner, C. *Agricultural and Forest Meteorology* **165**, 73–81 (2012). 10.1016/j.agrformet.2012.06.001.

[27] Chuine, I. *Philosophical Transactions of the Royal Society B: Biological Sciences* **365**(1555), 3149–3160 Sep (2010). 10.1098/rstb.2010.0142.

[28] Zohner, C. M., Benito, B. M., Svenning, J.-C., and Renner, S. S. *Nature Climate Change* **6**(12), 1120–1123 Oct (2016). 10.1038/nclimate3138.

[29] Wu, F., Gao, Y., Yang, W., Sui, N., and Zhu, J. *Front Plant Sci* **13**, 821563 (2022). 10.3389/fpls.2022.821563.

[30] Meng, L., Mao, J., Zhou, Y., Richardson, A. D., Lee, X., Thornton, P. E., Ricciuto, D. M., Li, X., Dai, Y., Shi, X., and et al. *Proceedings of the National Academy of Sciences* **117**(8), 4228–4233 Feb (2020). 10.1073/pnas.1911117117.

[31] Grimm, N. B., Faeth, S. H., Golubiewski, N. E., Redman, C. L., Wu, J., Bai, X., and Briggs, J. M. *Science* **319**(5864), 756–760 (2008). 10.1126/science.1150195.

[32] 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. *Journal of environmental management* **92**(3), 331–362 (2011).

[33] Zohner, C. M. and Renner, S. S. *Ecology Letters* **17**(8), 1016–1025 (2014). https://doi.org/10.1111/ele.12308.

[34] Ettinger, A. K., Gee, S., and Wolkovich, E. M. *Am J Bot* **105**(10), 1771–1780 Oct (2018). 10.1002/ajb2.1174.

[35] Primack, R. B. and Miller-Rushing, A. J. *New Phytologist* **182**(2), 303–313 (2009). https://doi.org/10.1111/j.1469-8137.2009.02800.x.

[36] McKown, A. D., Guy, R. D., Klápště, J., Geraldes, A., Friedmann, M., Cronk, Q. C. B., El-Kassaby, Y. A., Mansfield, S. D., and Douglas, C. J. *New Phytologist* **201**(4), 1263–1276 Nov (2013). 10.1111/nph.12601.

[37] Aitken, S. N. and Bemmels, J. B. *Evolutionary Applications* **9**(1), 271–290 Aug (2015). 10.1111/eva.12293.

[38] Vico, G., Karacic, A., Adler, A., Richards, T., and Weih, M. *BioEnergy Research* **14**(2), 445–459 (2021). 10.1007/s12155-021-10249-5.

[39] Scotti-Saintagne, C., Bodénès, C., Barreneche, T., Bertocchi, E., Plomion, C., and Kremer, A. *Theoretical and Applied Genetics* **109**(8), 1648–1659 (2004).

[40] Cuervo-Alarcon, L., Arend, M., Müller, M., Sperisen, C., Finkeldey, R., and Krutovsky, K. V. *Tree Genetics & Genomes* **14**(6), 84 (2018). 10.1007/s11295-018-1297-2.

[41] Saikkonen, K., Taulavuori, K., Hyvönen, T., Gundel, P. E., Hamilton, C. E., Vänninen, I., Nissinen, A., and Helander, M. *Nature Climate Change* **2**(4), 239 (2012).

[42] Way, D. A. and Montgomery, R. A. *Plant, Cell & Environment* **38**(9), 1725–1736 (2015). 10.1111/pce.12431.

[43] Gauzere, J., Delzon, S., Davi, H., Bonhomme, M., Garcia de Cortazar-Atauri, I., and Chuine, I. *Agricultural and Forest Meteorology* **244-255**, 9–20 (2017).

[44] Meng, L., Zhou, Y., Román, M. O., Stokes, E. C., Wang, Z., Asrar, G. R., Mao, J., Richardson, A. D., Gu, L., and Wang, Y. *PNAS Nexus* **1**(2) 04 (2022). pgac046. 10.1093/pnasnexus/pgac046.

[45] Dosmann, M. S. *The Botanical Review* **72**(3), 207–234 (2006). 10.1663/0006-8101(2006)72[207:RITGAT]2.0.CO;2.

[46] 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., and Svenning, J.-C. *Global Change Biology* **19**(5), 1470–1481 (2013). https://doi.org/10.1111/gcb.12129.

[47] de Rességuier, L., Mary, S., Le Roux, R., Petitjean, T., Quénol, H., and van Leeuwen, C. *Frontiers in Plant Science* **11**, 515 (2020). 10.3389/fpls.2020.00515.

[48] Lembrechts, J. J., Nijs, I., and Lenoir, J. *Ecography* **42**(7), 1267–1279 (2019). https://doi.org/10.1111/ecog.03947.

[49] Laigle, I., Carlson, B. Z., Delestrade, A., Bison, M., Van Reeth, C., and Yoccoz, N. G. *Frontiers in Earth Science* **10**, 10.3389/feart.2022.912048 (2022).

[50] Schwartz, M. D., Hanes, J. M., and Liang, L. *Agricultural and Forest Meteorology* **169**, 136–147 (2013). https://doi.org/10.1016/j.agrformet.2012.10.014.

[51] Whiteman, C. D., Hubbe, J. M., and Shaw, W. J. *Journal of Atmospheric and Oceanic Technology* **17**(1), 77 – 81 (01 Jan. 2000). 10.1175/1520-0426(2000)017<0077:EOAITD>2.0.CO;2.

[52] Hunter, A. F. and Lechowicz, M. J. **29**(3), 597–604 2020/10/20/ (1992). 10.2307/2404467.

[53] Finn, G., Straszewski, A., and Peterson, V. *Annals of Applied Biology* **151**(1), 127–131 Aug (2007). 10.1111/j.1744-7348.2007.00159.x.

[54] Denny, E. G., Gerst, K. L., Miller-Rushing, A. J., Tierney, G. L., Crimmins, T. M., Enquist, C. A. F., Guertin, P., Rosemartin, A. H., Schwartz, M. D., Thomas, K. A., and Weltzin, J. F. *International Journal of Biometeorology* **58**(4), 591–601 (2014). 10.1007/s00484-014-0789-5.

[55] O'Keefe, J. (2014).

[56] Stan Development Team. (2019).

[57] R Development Core Team. *R Foundation for Statistical Computing, Vienna, Austria* (2017).

[58] Asse, D., Chuine, I., Vitasse, Y., Yoccoz, N. G., Delpierre, N., Badeau, V., Delestrade, A., and Randin, C. F. *Agricultural and Forest Meteorology* **252**, 220–230 (2018). https://doi.org/10.1016/j.agrformet.2018.01.030.

[59] Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., and Rubin, D. B. *Bayesian Data Analysis*. CRC Press, New York, 3rd edition, (2014).

[60] Carpenter, B., Gelman, A., Hoffman, M., Lee, D., Goodrich, B., Betancourt, M., Brubaker, M. A., Guo, J., Li, P., and Allen, R. *Journal of Statistical Software* **76**(1), 10.18637/jss.v076.i01 (2017).

[61] Chang, W., Cheng, J., Allaire, J., Sievert, C., Schloerke, B., Xie, Y., Allen, J., McPherson, J., Dipert, A., and Borges, B. https://CRAN.R-project.org/package=shiny, (2021).

[62] Basler, D. *Agricultural and Forest Meteorology* **217**, 10 – 21 (2016). https://doi.org/10.1016/j.agrformet.2015.11.007.

[63] Jochner, S. and Menzel, A. *Environmental Pollution* **203**, 250–261 (2015). https://doi.org/10.1016/j.envpol.2015.01.003.

[64] Ettinger, A. K., Chamberlain, C. J., Morales-Castilla, I., Buonaiuto, D. M., Flynn, D. F. B., Savas, T., Samaha, J. A., and Wolkovich, E. M. *Nature Climate Change* **10**(12), 1137–U119 (2020).

[65] Luedeling, E., Girvetz, E. H., Semenov, M. A., and Brown, P. H. *PLOS ONE* **6**(5), 1–13 05 (2011). 10.1371/journal.pone.0020155.

[66] Fu, Y. H., Zhao, H., Piao, S., Peaucelle, M., Peng, S., Zhou, G., Ciais, P., Huang, M., Menzel, A., Peñuelas, J., and et al. *Nature* **526**(7571), 104–107 Sep (2015). 10.1038/nature15402.

[67] Baumgarten, F., Zohner, C. M., Gessler, A., and Vitasse, Y. *New Phytologist* **230**(4), 1366–1377 (2021). https://doi.org/10.1111/nph.17270.

[68] Zhang, R., Lin, J., Wang, F., Shen, S., Wang, X., Rao, Y., Wu, J., and Hänninen, H. *Agricultural and Forest Meteorology* **298-299**, 108296 (2021). https://doi.org/10.1016/j.agrformet.2020.108296.

[69] Stabler, L. B., Martin, C. A., and Brazel, A. J. *Urban Forestry Urban Greening* **3**(3), 137–147 (2005). https://doi.org/10.1016/j.ufug.2004.11.001.

[70] Erell, E., Pearlmutter, D., and Williamson, T. *Urban microclimate: designing the spaces between buildings*. Routledge, (2012).

[71] Dimoudi, A., Kantzioura, A., Zoras, S., Pallas, C., and Kosmopoulos, P. *Energy and Buildings* **64**, 1–9 (2013). https://doi.org/10.1016/j.enbuild.2013.04.014.

[72] Qu, M., Wan, J., and Hao, X. *Weather and Climate Extremes* **4**, 86–95 (2014). https://doi.org/10.1016/j.wace.2014.05.002.

[73] da Cunha, A. R. *Environmental Monitoring and Assessment* **187**(5), 236 (2015). 10.1007/s10661-015-4458-x.

[74] Vitasse, Y., Baumgarten, F., Zohner, C. M., Kaewthongrach, R., Fu, Y. H., Walde, M. G., and Moser, B. *New Phytologist* **232**(2), 537–550 (2021). https://doi.org/10.1111/nph.17606.

Tables and Figures

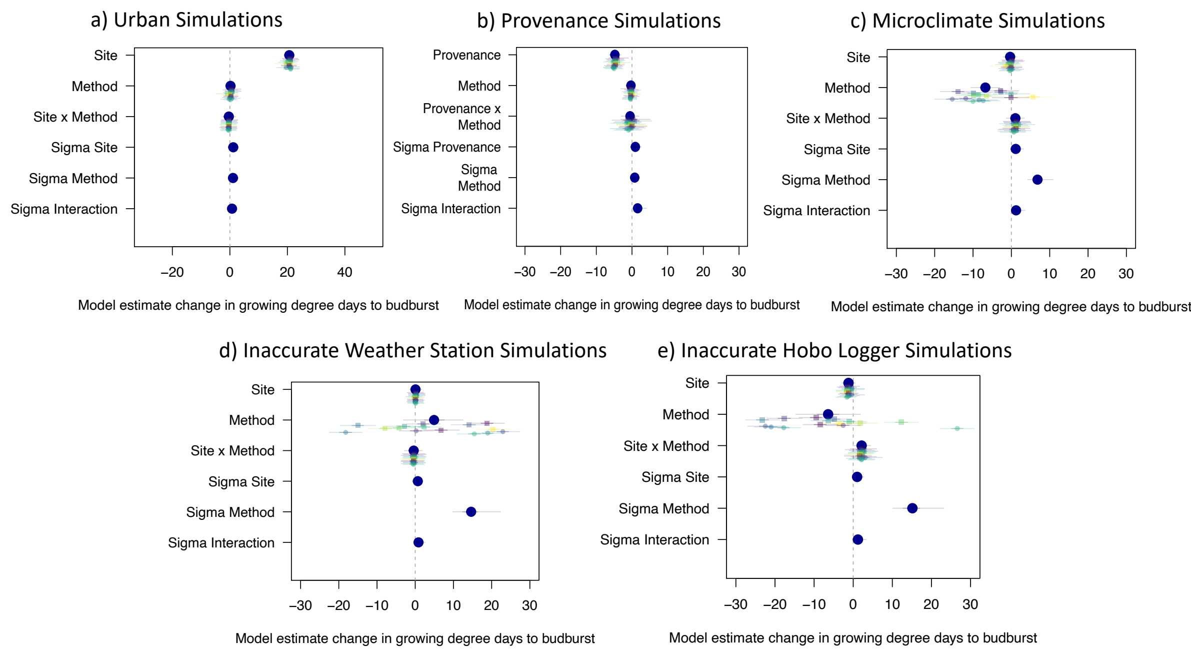

Figure 1: Simulations: we show (a) urban sites requiring more GDDs, (b) more northern provenance latitudes requiring fewer GDDs, (c) microclimate effects, (d) less accurate weather station data and (e) less accurate HOBO logger data, which looks similar to (c). We show the effects of site (urban versus 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.

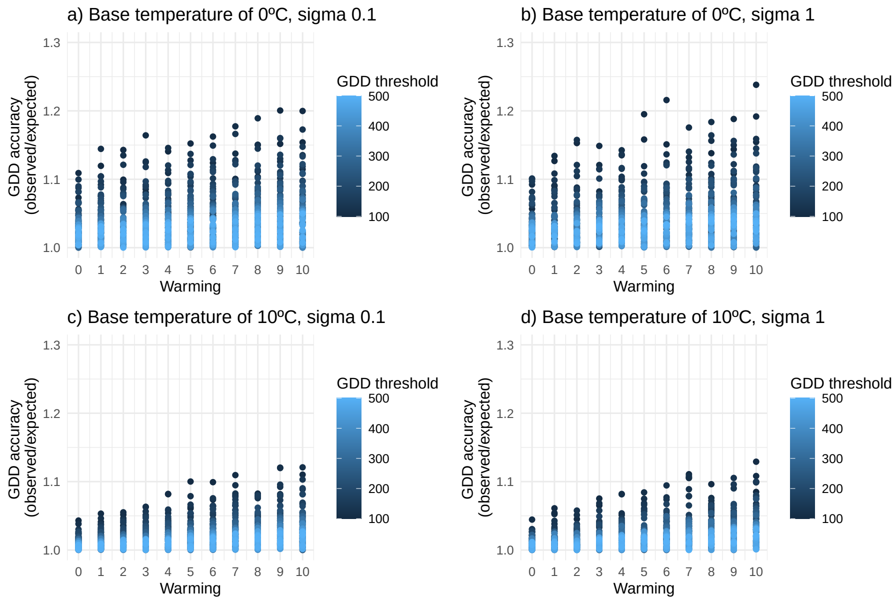

Figure 2: Using simulated data, we show how GDD measurement accuracy changes with warming (i.e., from 0°C to 10°C) using a base temperature of (a) 0°C and a sigma of 0.1°C, (b) 0°C and a sigma of 1°C, (c) 10°C and a sigma of 0.1°C and (d) 10°C and a sigma of 1°C. GDD accuracy is measured as the observed 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.

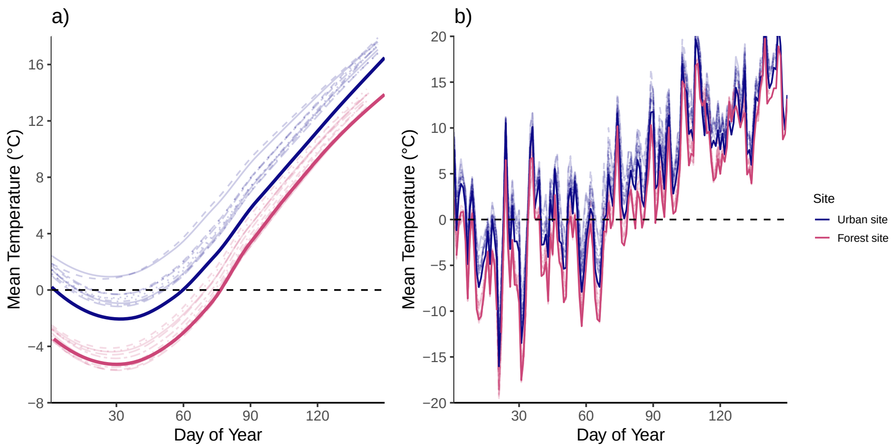

Figure 3: Here we show a breakdown of the climate data across the two sites with darker lines representing weather station data and the lighter, more transparent lines of varying line types representing the HOBO loggers: a) a series of smoothing splines of mean temperature with 90% uncertainty interval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

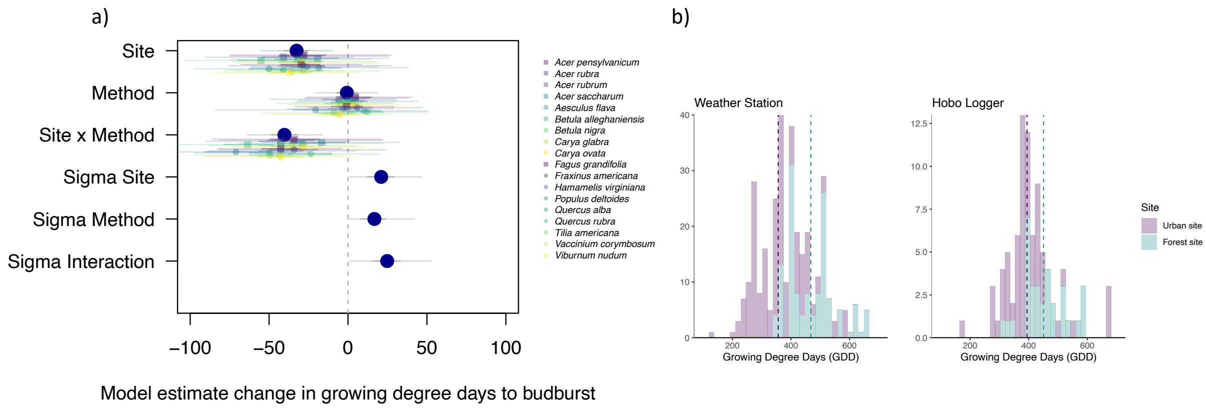

Figure 4: Empirical Data: we show (a) the main effects and variance (sigma) of site (urban versus rural) and climate data method (weather station versus HOBO loggers), as well as their interaction, on GDDs until budburst. The intercept represents the HOBO logger data for the rural forested site. More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

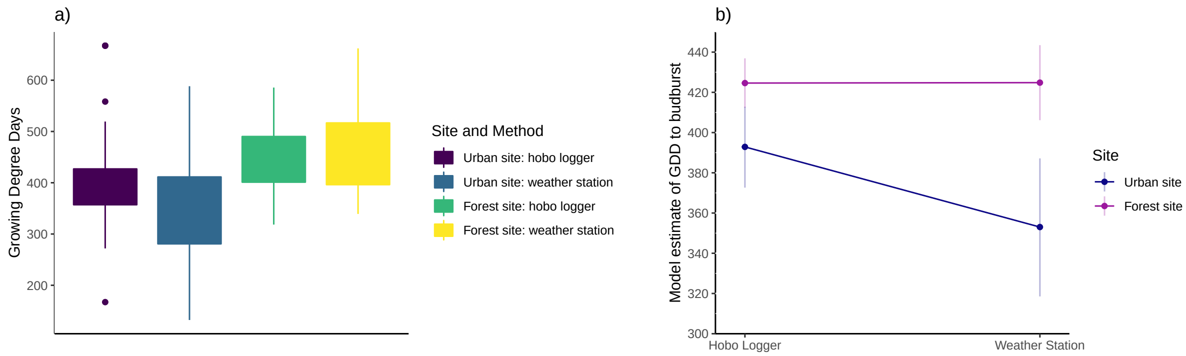

Figure 5: We show estimated effects, from a Bayesian hierarchical model, of site (urban arboretum site versus forested rural site) by climate data method (weather station data versus HOBO logger data) on GDDs until budburst (a) as a boxplot across each method and site combination using raw data and (b) using model output to show the mean estimates for each site and method with 50% uncertainty intervals shown as error bars. Modeled estimates suggest there was a difference of approximately 30 GDDs between the urban and rural sites using the hobo loggers, whereas the difference was approximately 70 GDDs using weather station data.

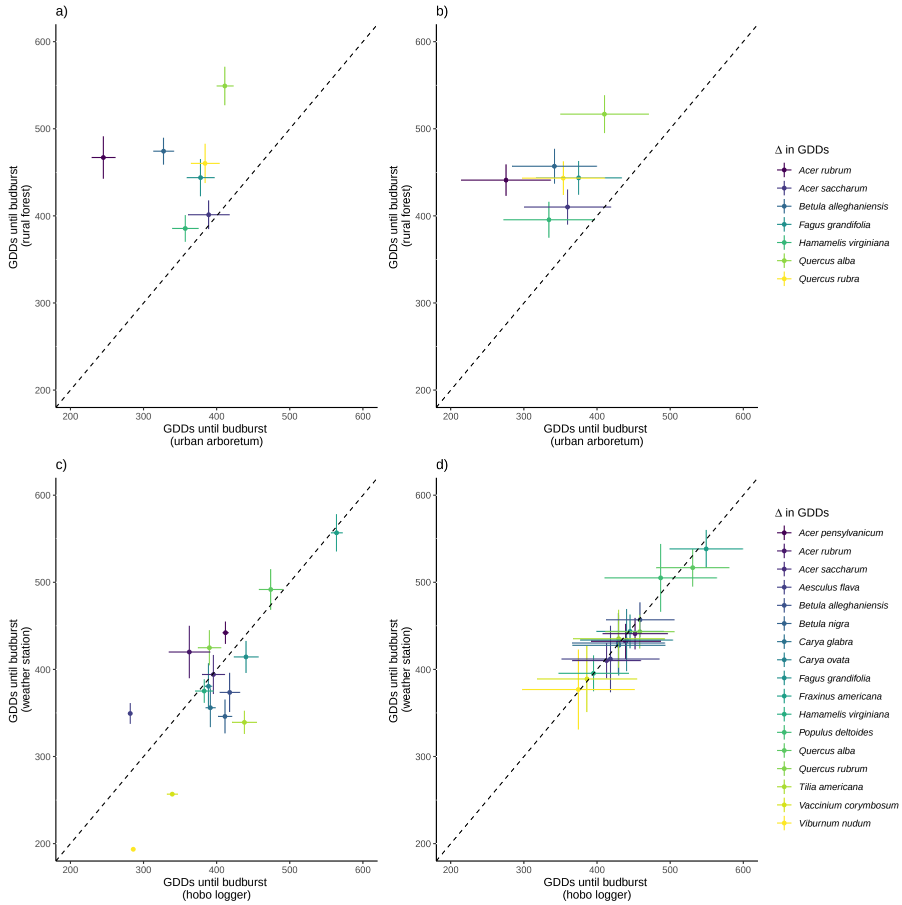

Figure 6: Empirical Data: Using real data, we show (a) the effects of site (urban versus rural) on GDDs until budburst across all species used in the study. We also show model output for the same relationship in (b). In panel (c), we show the effects of climate data method (weather station versus HOBO logger) on GDDs until budburst across all species and see the model output for the same relationship in (d). Using empirical data, we see that individuals at the rural site generally require more GDDs until budburst than trees and there is large variation in GDD requirement between the two climate data methods (see 1:1 dashed line).