

## Supplemental information -- Model background

Here we provide background on the hierarchical model to determine how continuously varying temperature affects phenology. First we discuss how our approach relates to  $\alpha$ - and  $\beta$ -sensitivity estimated in current studies and why they produce different estimates. We then describe methods used in this study to evaluate  $\gamma$ -sensitivity.

### Current methods

Disagreement between experiments and observational studies begins with distributional properties of their estimates. For both measures, we follow standard practice in reporting sensitivities as positive values when the coefficient is negative, indicating that increasing temperature decreases time to BB. For  $\beta$ -sensitivity, the distribution is approximately normal,

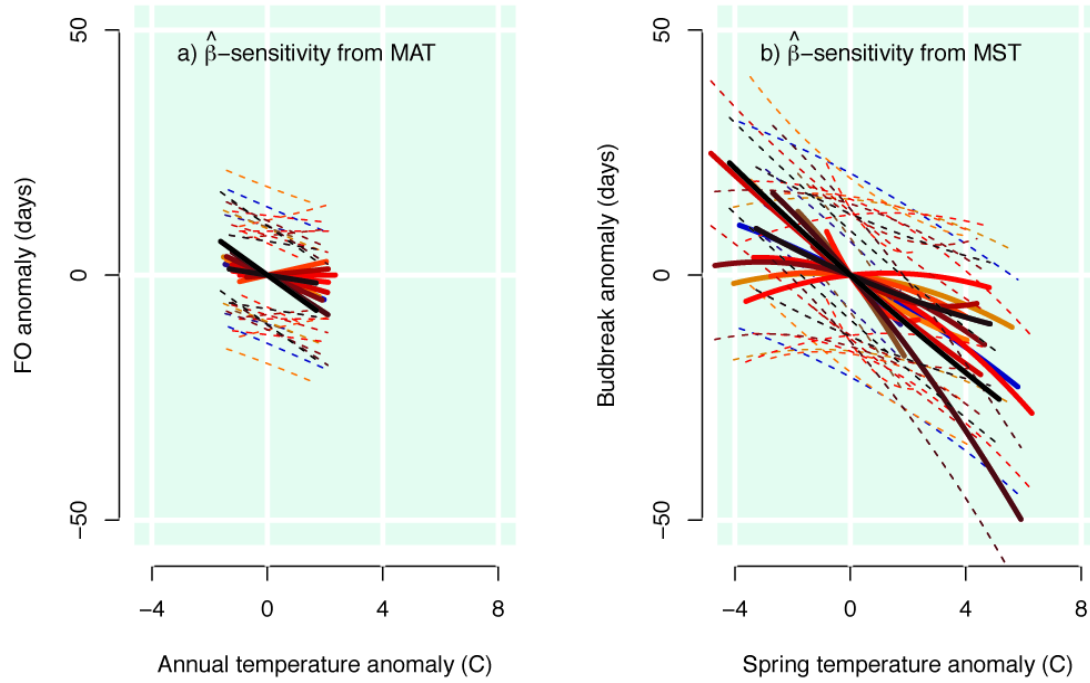
$$\hat{\beta} \sim N(\mu_\beta, \nu_\beta) \quad \text{S1}$$

with mean  $\mu_\beta = \text{cov}(FO_{iy}, MAT_{iy}) / \text{var}(MAT_{iy})$ , variance  $\nu_\beta = \frac{\hat{\sigma}^2}{\text{var}(MAT_{iy})}$ , and mean residual

standard error  $\hat{\sigma}^2 \approx (NY - 1)^{-1} \sum_{i,y} (FO_{iy} - \hat{\beta} MAT_{iy})^2$  for  $N$  observations over  $Y$  years, which could include individual plants in each observation year (estimates do not typically incorporate dependence structure involving individuals and years).

The estimate from eqn S1 is subject to at least two types of errors that are not accommodated by regression. First, the regression model assumes dates and temperatures are known, and it admits Gaussian error in the relationship between them. The true date of budbreak is related to the first occasion as  $BB_{iy} \leq FO_{iy}$ , i.e., not Gaussian, and not centered on  $BB_{iy}$ . Second, the predictor  $MAT_{iy}$  is not the variable that controls budbreak, the problem long recognized by physiologists. By comparing estimates based on  $BB_{iy}$  vs  $FO_{iy}$  and on  $MAT$  vs  $MST$ , we show these errors have large impact on results, and that the relationship between them is not consistent (unshaded bars in Fig. 4 of the main text).

The impact of temperature aggregation on sensitivity estimates depends on how the aggregation is done. In observational studies temperature data are obtained from the closest weather station or from an interpolated grid. In Wolkovich et al. (2012) mean annual temperature for a year  $y$ ,  $MAT_y$ , come from the Global Historical Climatology Network and the European Climate Assessment. If the variable  $MAT_y$  was simply a noisy version of the true temperature that controls  $BB$ , then  $\hat{\beta}$  would underestimate true sensitivity. This is termed the error-in-variables problem, where the predictor (temperature) has error in addition to that contributed by  $\varepsilon_{iy}$ . However, if  $MAT_y$  values obtained from spatio-temporally aggregated data fundamentally differ from the important temperature variation experienced by plants, it is not possible to generalize about its effects on sensitivity estimates.



**Figure S1. Comparison of sensitivity coefficients applied in observational studies, defined as advance in BBD per °C.** Regression based on the narrow range of highly aggregated MAT (a) compared with the same model based on MST (b). Quadratic terms are included where significant. The coefficients, slopes of these regressions at (0, 0), take a range of values depending on how the aggregation is done. Each color represents a different species.

Estimates based on experiments have a different distribution. For  $\alpha$ -sensitivity, the distribution is approximately

$$\hat{\alpha} \sim N(\mu_{\alpha}, v_{\alpha}) \quad \text{S2}$$

with mean  $\mu_{\alpha} = \frac{\overline{FO_e} - \overline{FO_c}}{T_e - T_c}$ , variance  $v_{\alpha} = \frac{1}{R(T_e - T_c)^2} [\text{var}(\overline{FO_e}) + \text{var}(\overline{FO_c})]$ , where  $R$  is the

number of replicates contributing to means and variances for  $FO$ , such as the number of warming chambers<sup>1</sup>. The standard error of the estimate (the square root of  $v_{\alpha}$ ) can be large simply because replication is at the chamber level (or study level in meta-analysis), rather than the individual level in eqn S1. The small  $R$  (note  $R \ll N$ ) results in standard errors so large (Table S2) that they could not be plotted in Figure 4 of the main text. (Mean estimates for four years are shown as crosses in Figure 4a of the main text.) Of course, if meta-analysis considers only a single value for a study, there is no standard error. Large error in these estimates makes them of limited use for predicting responses to future climate change.

<sup>1</sup> In this case, the denominator is fixed by design (temperatures are controlled). A more general case where the denominator is random (e.g., Stuart and Ord 1994) additionally depends on the covariance between predictor and response.

Thus, both estimates suffer from aggregating temperature variation into a single number for a year (observational studies) or a treatment replicate (experiments). The two methods for analysis are not directly comparable, despite appearing similar in dimension.

## 2.2 The continuous development model

The continuous development model (CDM) is a type of state-space model, which allows for a hierarchical structure of [data|process, parameters][process|parameters][parameters] (Calder et al. 2003). There is a discrete ordinal scale for phenology stages (Table 2). Phenological development is initiated by warming temperatures, and it responds to temperature variation on a continuous basis. But it is observed only in terms of the six discrete ordinal classes.

Development depends on continuously changing temperatures, the current developmental state, and the current state of the individual, including genetic variation associated with seed source. Although observations are discrete the state changes themselves are not, nor is the precise timing of changes between them. The model for process and observations is summarized in this section.

**Table S1** Summary of fits for  $\beta$  in Figure 4 of the main text. Standard regression is used to match standard analysis in the literature.

Species	Estimate	Std.Error	t value	Pr(>  t  )
<b>From MST</b>				
list	-2.00	0.410	-4.90	1.6e-06
beal	-6.80	0.150	-45.0	2.2e-183
bepa	-7.30	0.180	-41.0	1.2e-97
litu	-0.62	0.190	-3.20	1.4e-03
magr	-2.90	1.10	-2.60	9.9e-03
pipa	-1.70	0.790	-2.10	3.8e-02
pire	-5.30	0.400	-13.0	1.6e-23
pita	-2.90	0.210	-13.0	1.6e-33
acru	-1.20	0.086	-14.0	1.3e-45
acsa	-2.30	0.330	-7.10	2.2e-11
fram	-0.48	0.440	-1.10	2.7e-01
nysy	-0.35	0.380	-0.92	3.6e-01
pist	-4.80	0.270	-18.0	2.2e-49
qual	-0.99	0.070	-14.0	9.5e-43
quru	-3.50	0.230	-16.0	2.8e-42
list	-2.50	0.82	-3.10	2.2e-03
beal	-1.50	0.55	-2.70	6.6e-03
bepa	-2.20	0.97	-2.30	2.3e-02
<b>From MAT</b>				
litu	-2.40	0.37	-6.60	1.2e-10
magr	-2.20	1.80	-1.30	2.1e-01
pipa	1.40	1.30	1.10	2.9e-01
pire	-2.70	0.82	-3.40	1.1e-03
pita	-0.69	0.51	-1.40	1.8e-01
acru	0.0082	0.19	0.044	9.6e-01
acsa	-1.70	0.54	-3.10	2.0e-03
fram	0.60	0.61	0.990	3.3e-01
nysy	-2.70	0.67	-4.0	9.1e-05
pist	-3.90	0.66	-5.80	1.5e-08

qual	-0.95	0.14	-6.900	5.5e-12
quru	-4.20	0.50	-8.500	6.2e-16

### Model stages

There is an observed discrete stage  $S_{iy,t}$   $S_{iy,t} \in \{1, \dots, K\}$  for individual  $i = 1, \dots, n$  at sample times  $\{t\}_i$  during year  $y$ . The  $K = 6$  observed stages for our study, ranging from dormant to fully expanded leaves, are listed in Table 1. There is underlying true stage that is also discrete  $s_{iy,t} \in \{1, \dots, K\}$ . There is multinomial observation error describing the probability that an individual is observed to be in stage  $S_{iy,t}$ , when its true stage is  $s_{iy,t}$ ,

$$p(S_{iy,t} | s_{iy,t} = k) \sim \text{multinom}(1, \pi_k) \quad \text{S3}$$

where  $\pi_k = [\pi_{k1}, \dots, \pi_{k7}]$  is the vector of probabilities that an individual in state  $k$  is recorded as any of  $K$  states. Observation error allows for the fact that states are not precisely discrete, but are recorded as such. The  $K \times K$  matrix  $\pi$  is a concatenation of the six  $\pi_k$  vectors, each of length  $K$ .

**Table S2** Estimates of  $\alpha$  from Figure 2 of the text with point estimates and standard errors from eqn S2. Not all species have large enough populations for estimates at both sites.

	DF +3		DF+5		HF+3		HF+5 treatment	
	treatment		treatment		treatment			
	Estimate	se	Estimate	se	Estimate	se	Estimate	se
list	0.062	1.0	-3.40	1.20	-6.60	-	-4.30	-
beal	-	-	-	-	-0.88	0.79	-0.450	2.80
bepa	-	-	-	-	-2.0	0.84	-1.30	0.62
litu	-2.50	1.50	-1.50	0.46	-2.30	-	-2.40	-
magr	-3.50	1.60	-2.70	0.68	-	-	-	-
pipa	0.570	1.60	-1.90	0.22	-	-	-	-
pire	-	-	-	-	-2.20	0.49	-1.70	0.20
pita	0.330	1.0	0.051	0.47	-	-	-	-
acru	-0.140	0.65	0.560	0.49	-1.20	0.18	-0.720	1.80
acsa	-3.20	-	0.760	6.10	-0.81	-	-2.10	-
fram	-1.30	-	-0.430	-	-0.78	-	-0.760	-
nysy	-4.40	0.54	-1.90	0.53	-4.60	-	-2.90	-
pist	-2.10	6.40	-1.90	1.50	1.10	0.24	0.028	1.0
qual	-1.20	0.59	-0.850	0.28	-0.72	1.30	-1.70	0.98
quru	-2.80	1.0	-1.90	1.20	-1.30	0.87	-0.480	2.10

The discrete stages  $s_{iy,t}$  arise as observable changes that result from continuous development. This latent continuous developmental state  $h_{iy}(t)$  is discretized as  $h_{iy,t}$ , where  $t$  is taken to be the time index of the model. There are corresponding probabilities that the individual will be in each of the observable states  $s_{iy,t}$ . The developmental state  $h_{iy,t}$  increases from 0 at onset to 100 at the time of bud break. The range of the latent scale  $h$  is not important, because we are ultimately interested in how environmental variation affects the probability of change in observable states--we marginalize over  $h$  to move from environment to probability of observable states. In other words,  $h$  is a device that allows for the fact that development is essentially continuous, responding instantaneously to the environment experienced by the individual  $x_{iy,t}$ . We use the model

$$\begin{aligned} dh_{iy,t} &= \exp(x_{iy,t}b + \varepsilon_{iy,t}) \left(1 - h_{iy,t}/h_{\max}\right) dt \times I(T_{iy,t} > T') \\ \varepsilon_{iy,t} &\sim N(0, \sigma^2) \end{aligned} \quad \text{S4}$$

where  $dt = 2$  days. The indicator function  $I(\cdot)$  takes the value of 1 when temperature is greater than a threshold  $T' = 0^\circ\text{C}$ , and zero otherwise. This simple formulation was selected to have readily interpretable parameters and to capture the basic features of development toward a final state approached at asymptote  $h_{\max} = 100$ . Incremental development is non-negative (it can approach zero, but not regress). Covariates that accelerate development when positive are associated with positive values in parameter vector  $b$ , and vice versa. Input variables occupy a design vector

$$x_{iy,t} = \begin{bmatrix} 1 & T_{iy,t} & G_i & T_{iy,t} \times G_i & C_{iy} \end{bmatrix} \quad \text{S5}$$

where  $T_{iy,t}$  is the average temperature for day  $t$  experienced by individual  $i$ ,  $G_i \in \{0,1\}$  is a binary indicator for southern (0) vs northern (1) seed source,  $T_{iy,t} \times G_i$  is the interaction,

$$C_{iy} = n_{iy} (T'_c - \bar{T}_i)$$

are chilling units, and  $n_{iy} = \sum_i I(T'_c - T_{iy,t})$  is the number of days below the chilling threshold  $T'_c = 0^\circ\text{C}$ .

The discrete stages  $s_{iy,t}$  comprise an ordinal scale that is linked probabilistically to  $h_{iy,t}$ . As  $h_{iy,t}$  increases over time, individuals transition through the discrete states from  $s_{iy,t} = 1$  when dormant through  $s_{iy,t} = 6$  when leaves are fully expanded. The connection between  $h_{iy,t}$  and  $s_{iy,t}$  is established by a multinomial logit,

$$s_{iy,t} \sim \text{multinom}(1, \theta_{iy,t}) \quad \text{S6}$$

where

$$\begin{aligned} \theta_{iy,t,k} &= P_{iy,t,k} - \sum_{j=1}^{k-1} P_{iy,t,j} \\ \ln\left(\frac{P_{iy,t,k}}{1 - P_{iy,t,k}}\right) &= c_{k0} + c_{k1} h_{iy,t} \end{aligned} \quad \text{S7}$$

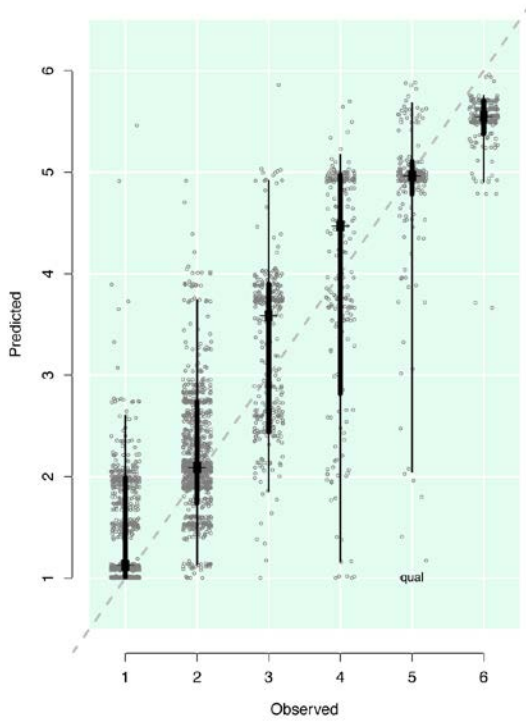
$\theta_{iy,t,k}$  is the probability that individual  $i$  at time  $y,t$  is in stage  $k$ .  $P_{iy,t,k}$  is the probability that the individual has not yet progressed beyond state  $k$ . There is a  $(K-1) \times 2$  matrix of coefficients of

intercepts  $c_{k0}$  and slopes  $c_{k1}$  on the logit scale that determine the transitions between states. The ordinal states imply the relationships  $(c_{k0} < c_{(k+1)0}, c_{k1} < c_{(k+1)1})$ .

Prior distributions for the model are non-informative, with several exceptions. First, the observation errors are informative with prior

$$\pi_k \sim \text{Dirichlet}(r_k)$$

to admit departures from observations when there are reversals in observed states, i.e., an observation at one time implies an earlier stage than a previous time,  $S_{iy,t'} < S_{iy,t}$  for  $t' < t$ . All observational information enters through the  $S_{i,t}$ , so it is important to make them informative. Organized as a matrix, the  $K = 6$  prior vectors are:



**Figure S2.** Full season predictions of six states from day of the year 20 for all individuals and of *Q. alba*. Dots are individuals at 2-day intervals. Thick and thin lines span 68% and 95% of predicted values.

has mean value of 2, which is approximately the maximum change in  $h_{iy,t}$  for a time increment  $dt$ .

The prior distribution for the  $c$  parameters in eqn S7 is informative for transitions from the first stage and into the last stage, but otherwise weak. The difference between prior (dotted lines) and posterior (solid lines) in Figure 1 of the main text shows how the prior is updated by data.

		True state:					
		$r_1$	$r_2$	$r_3$	$r_4$	$r_5$	$r_6$
Observed state:							
1		$N$	0.1	$10^{-4}$	$10^{-4}$	$10^{-4}$	$10^{-4}$
2		0.1	$N$	0.1	$10^{-4}$	$10^{-4}$	$10^{-4}$
3		$10^{-4}$	0.1	$N$	0.1	$10^{-4}$	$10^{-4}$
4		$10^{-4}$	$10^{-4}$	0.1	$N$	0.1	$10^{-4}$
5		$10^{-4}$	$10^{-4}$	$10^{-4}$	0.1	$N$	0.1
6		$10^{-4}$	$10^{-4}$	$10^{-4}$	$10^{-4}$	0.1	$N$

where  $N$  is the number of observations. The prior distribution on  $b$  in eqn S4 is flat, but truncated at zero for temperature. Posterior distributions were well above zero (see Figure S2).

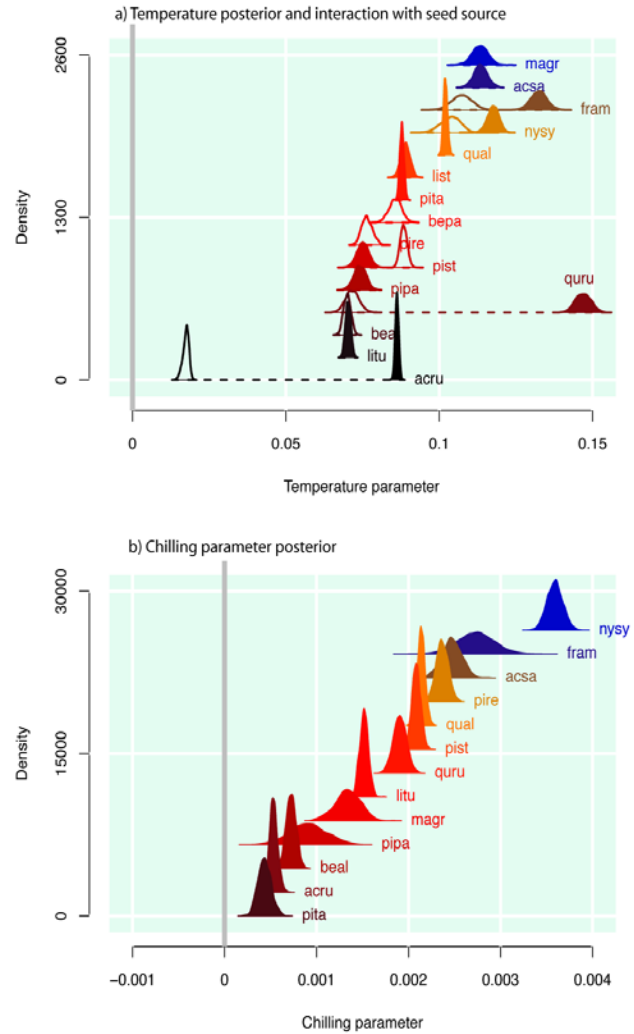
The prior distribution for the process variance  $\sigma^2 \sim IG(n, 2(n-1))$

Posterior simulation was done with Metropolis-within-Gibbs. The germination year was not included in the analysis to avoid possible environmental effects resulting from germination itself. The sample sizes in Table 2 refer to post-germination year individuals. Model implementation at time steps of  $dt = 1$  and  $dt = 2$  yielded the same results. There is a state space model for each individual, initialized at  $h_{iy,20} = 1$  on day 20 of each year, continuing through fully expanded leaves. Censuses at weekly intervals provide the  $S_{iy,t}$ . Convergence occurred within 20,000 iterations, which were discarded. Results reported here come from an additional 50,000 iterations. Diagnostics involve model checking through in-sample prediction. The model was evaluated by predicting the entire course of observations for all individuals starting from dormant state 1 through fully expanded state 6 (Fig. S2). This is a more rigorous standard than the one-step-ahead prediction often used to evaluate fitted time series models. There is large variability, but most 68% (1 standard deviation of the mean) and all 95% predictive intervals spanned the 1:1 line. Model selection was done with the deviance information criterion (DIC) for the variables in eqn S5. These are the models retained in Table 1. Example posterior distributions for temperature and chilling are shown in Figure S3.

#### *From the development scale to discrete observations*

$\gamma$ -sensitivity of development rate summarizes the effect of temperature on continuous progress toward budbreak at stage  $k = 3$ . We use the probability of state 3 (budbreak),  $P_3$ , rather than  $h$ , because  $h$  is not observed, whereas  $P_3$  is directly related to observed event, the probability of being in at least stage  $k$ . This changes at rate

$$\frac{\partial(1 - P_k)}{\partial t} = - \frac{\partial P_k}{\partial h} \frac{dh}{dt}$$



**Figure S3.** Posterior densities for temperature and chilling parameters (eqn S5). For species having a significant interaction between seed source and temperature, northern populations have unshaded densities in (a). Positive values accelerate development. For chilling, cold winters increase subsequent development rate.

From eqn S7, the first factor is logistic,

$$Q_k(h) = -\frac{\partial P_k}{\partial h} = \frac{-c_{(k-1)l} e^{c_{h,k-1}}}{(1 + e^{c_{h,k-1}})^2} \quad \text{S8}$$

For a benchmark comparable across species, we evaluate eqn S8 at the value of  $h$  where  $P_k = 1/2$ ,

$$Q_k(h)_{P_k=1/2} = \frac{-c_{(k-1)l}}{4}, \quad \text{S9}$$

spring temperature of 10°C, and the mean chilling units taken over all years of the study. The sensitivity to temperature  $x$  is expressed on this scale as a rate sensitivity,

$$\begin{aligned} \gamma_k &= \frac{\partial}{\partial T} \left( Q(h) \frac{dh}{dt} \right)_{P_k=1/2} = Q_k(h) \frac{\partial}{\partial T} \left( \frac{dh}{dt} \right)_{P_k=1/2} \\ &= \frac{-c_{(k-1)l}}{4} \left( b_T + \sum_{\{x'\}} x' b_{Tx'} \right) \frac{dh}{dt} \end{aligned} \quad \text{S10}$$

where  $x'$  is an input variable that interacts with temperature, e.g., seed source,  $\{x'\}$  is the set of all interactions, and  $dh/dt$  is approximated by dividing the right-hand side of eqn S4 by  $dt = 2$  days. The uncertainty in  $\hat{\gamma}_3$  estimates is obtained by Monte Carlo integration (Table S2). These are the values reported in Figure 4a of the main text.

For the timing of maximum sensitivity in Figure 5 of the main text we evaluate sensitivity continuously for each individual, conditional on the current state and temperature, i.e. eqn S4 and S8. These sensitivities increase with warming in spring to reach the maximum values by dates shown in Figure 2, and subsequently decline.

## References

- Augsburger CK, Bartlett EA (2003) Differences in leaf phenology between juvenile and adult trees in a temperate deciduous forest. *Tree Physiology* 23, 517-525.
- Calder, K. M. Lavine, P. Mueller, and J.S. Clark (2003) Incorporating multiple sources of stochasticity in population dynamic models. *Ecology* 84:1395-1402.
- Cleland, E., et al. (2007) Shifting plant phenology in response to global change. *Trends Ecol. Evol.* 22, 357–365.
- Linkosalo T, Häkkinen R, & Hänninen H. (2006) Models of the spring phenology of boreal and temperate trees: Is there something missing? *Tree Physiol.* 26, 1165-72.
- Stuart, A. and J.K. Ord. (1994) *Kendall's Advanced Theory of Statistics*, Sixth Edition. Edward Arnold, London.
- Wolkovich, E.M. et al. (2012) Warming experiments underpredict plant phenological responses to climate change. *Nature*, doi:10.1038/nature11014.