Modeling biological processes as stopped random walks

Jonathan Auerbach
jauerba@gmu.edu

EFI and Statistical Ecology Section Webinar

December 2, 2024

STATISTICS
George Mason University

Contents

STATISTICS
George Mason University

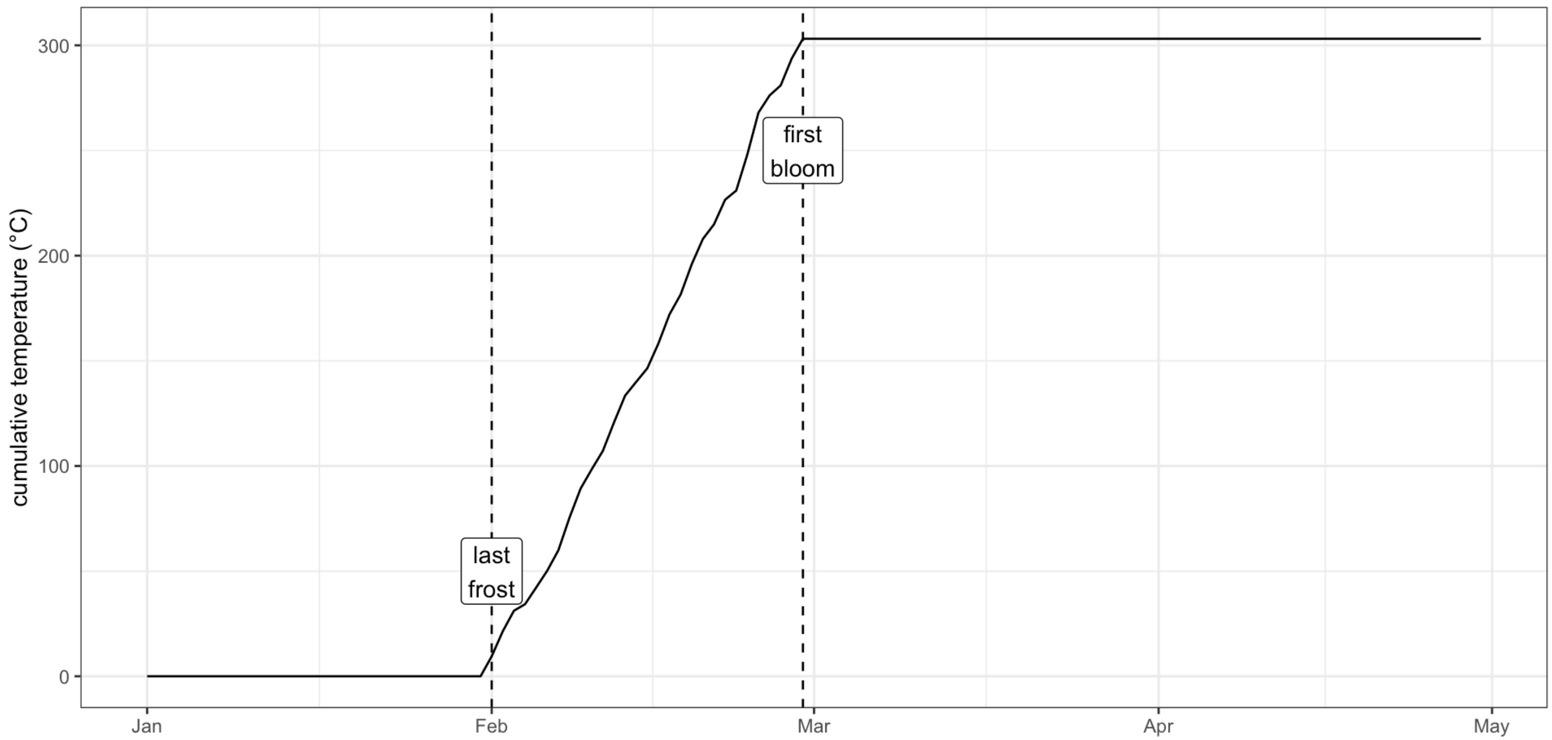
Stopped Random Walks

Many biological processes can be modeled as stopped random walks

STATISTICS
George Mason University

Simulation

▶ expand for R code



a simulation of the law of flowering plants: a flower first blooms in the spr
temperatures reach a threshold

STATISTICS
George Mason University

Two assumptions

Let $X_i > 0$ denote the force exerted on the plant on day i.

- Assume
  1. X_i have common mean μ (the ambient temperature).
  2. X_i are independent with common variance σ^2.

- Both assumptions can be relaxed.
  - e.g., 1. can be replaced with the assumption that X_i have mean μ_i, and $\sum \mu_i$ does not grow too quickly (regularly varying with index 1).

STATISTICS
George Mason University

CLT for stopped random walks

Let $S_a^i = \sum_{j=a}^{i} X_j$ denote the cumulative force from day a to i

Let n_γ denote the bloom date.

- Assume the plant blooms when S_a^i first passes γ,

$$n_\gamma = \min\{i : S_a^i \geq \gamma\}$$

When γ is large,

$$n_\gamma \mathrel{\dot\sim} \text{Normal}\left(\gamma/\mu + a,\ \gamma\sigma^2/\mu^3\right)$$

STATISTICS
George Mason University

Simulation ($a = 32, \gamma = 300$)

▶ expand for R code

bloom_date	mean_temp
2024-03-25	5
2024-02-28	10

▶ expand for R code

term	estimate	std.error	statistic	p.value
a	31.62	0.37	86.36	0
γ	297.13	7.09	41.91	[illegible]

STATISTICS
George Mason University

Three takeaways from CLT

Recall bloom date $n_\gamma \dot\sim \text{Normal}\left(\gamma/\mu + a,\ \gamma\sigma^2/\mu^3\right)$. This imples:

1. $\mathbb{E}[n_\gamma - a] \propto 1/\mu$ and variance $\text{Var}\left(n_\gamma\right) \propto 1/\mu^3$
2. γ depends on the temperature scale (°C, °F, etc.) and is only identified if μ is known
3. When a unknown, we can count from any $b > 0$ since for $\delta = a - b$,

$$n_\gamma + \delta \dot\sim \text{Normal}\left(\gamma/\mu + \delta,\ \gamma\sigma^2/\mu^3\right)$$

STATISTICS
George Mason University

Experimental Data

Experiment (Charrier et al. 2011)

- 30 walnut trees (*Juglans* sp.), 6 genotypes at 2 locations.
  - Stems sampled in November from each tree and cut in 7-cm-long pieces with only one bud.
  - Stems chilled (4°C) and then forced (warmed) at different temperatures (5, 10, 15, 20 and 25°C).

- We examine the relationship between:
  - forcetemp: temperature during forcing
  - response.time: (day of budburst) when 50% of buds unfolded (stage 15 of BBCH scale)

STATISTICS
George Mason University

Experiment (Charrier et al. 2011)

▶ expand for R code

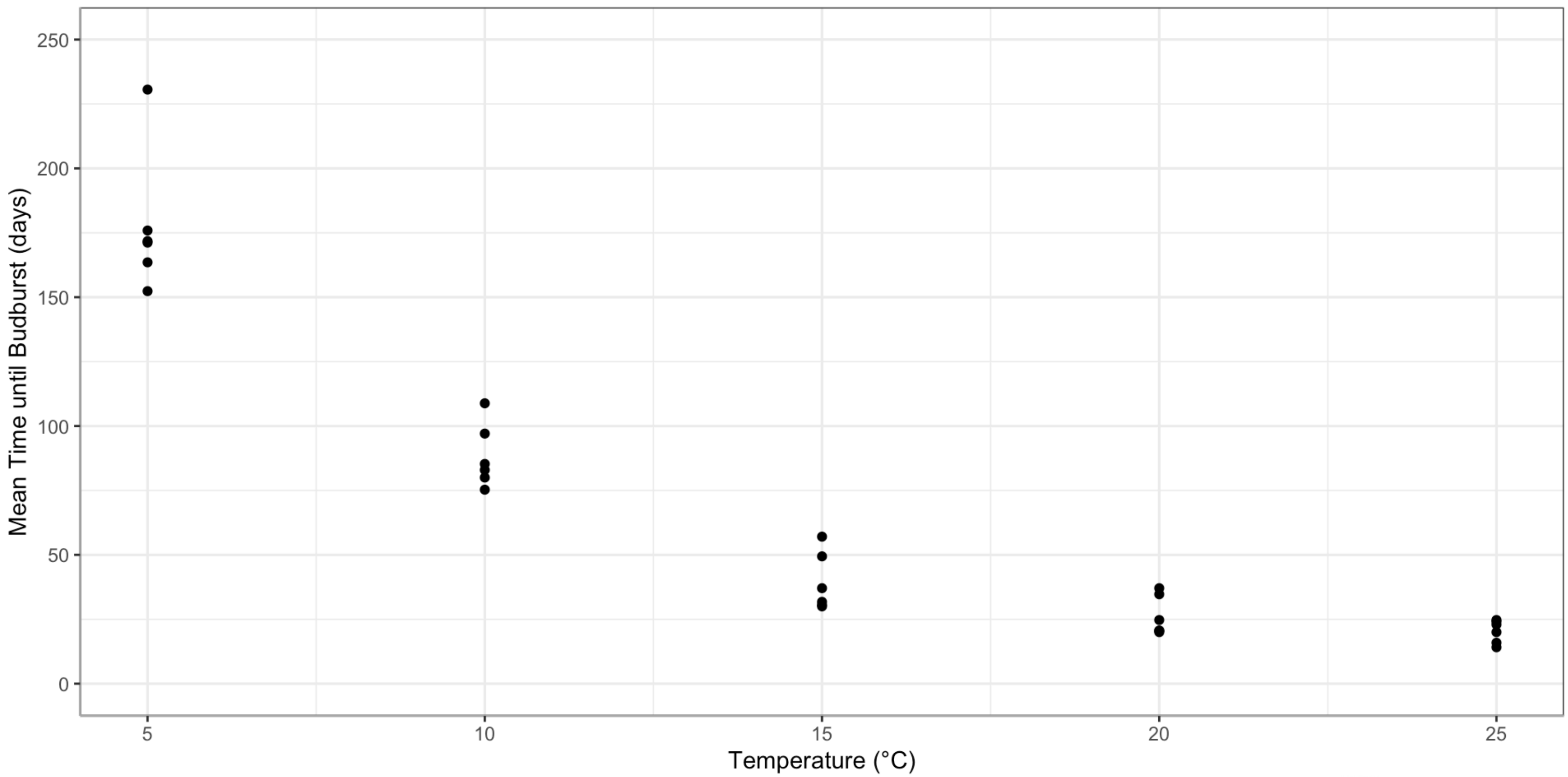
forcetemp	response.time
5	175.88
10	97.06
15	31.76
20	24.71
25	22.94

data from Charrier et al. (2011)

STATISTICS
George Mason University

Experiment (Charrier et al. 2011)

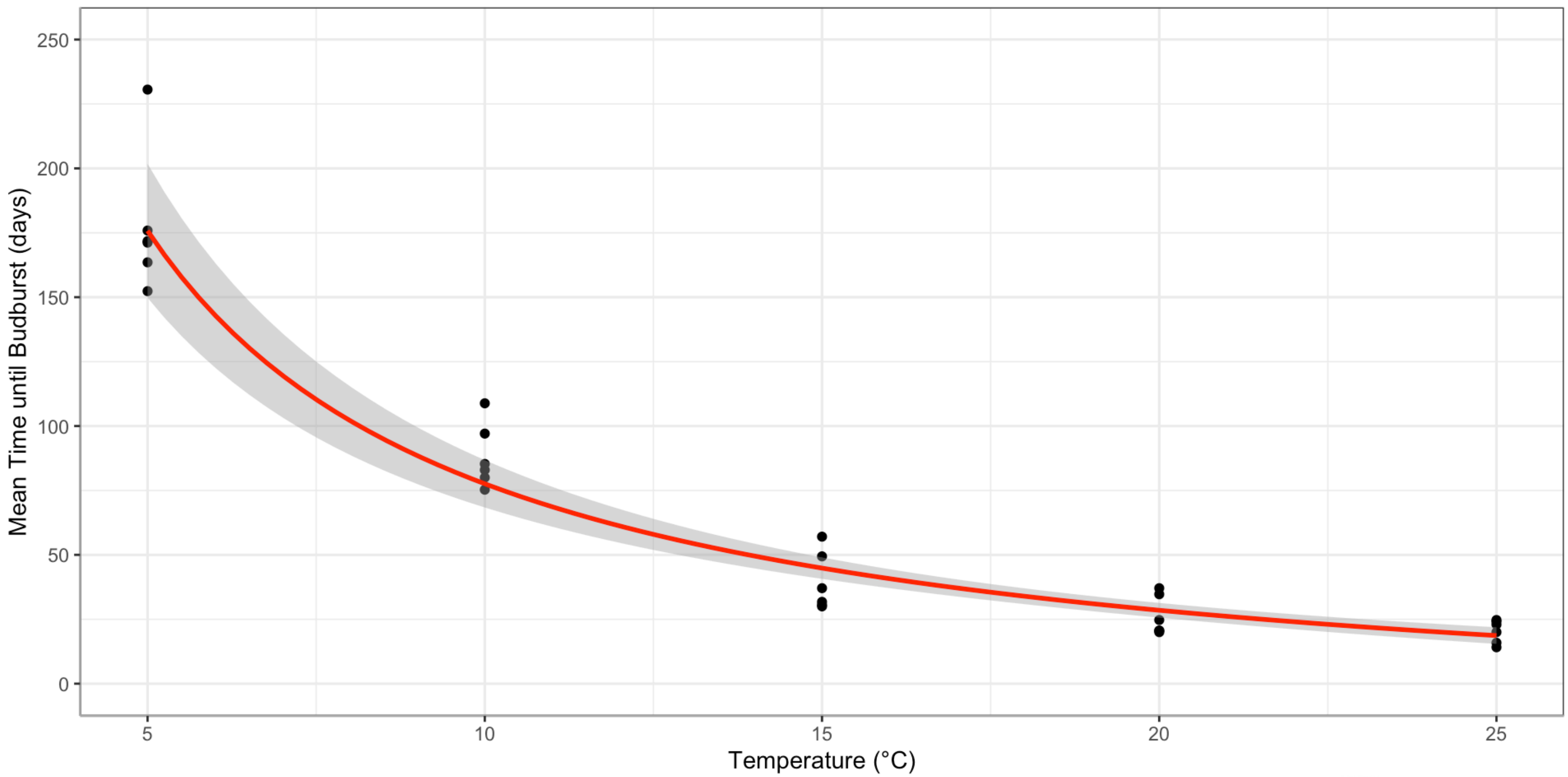
▶ expand for R code



data from Charrier et al. (2011)

Fit using lm

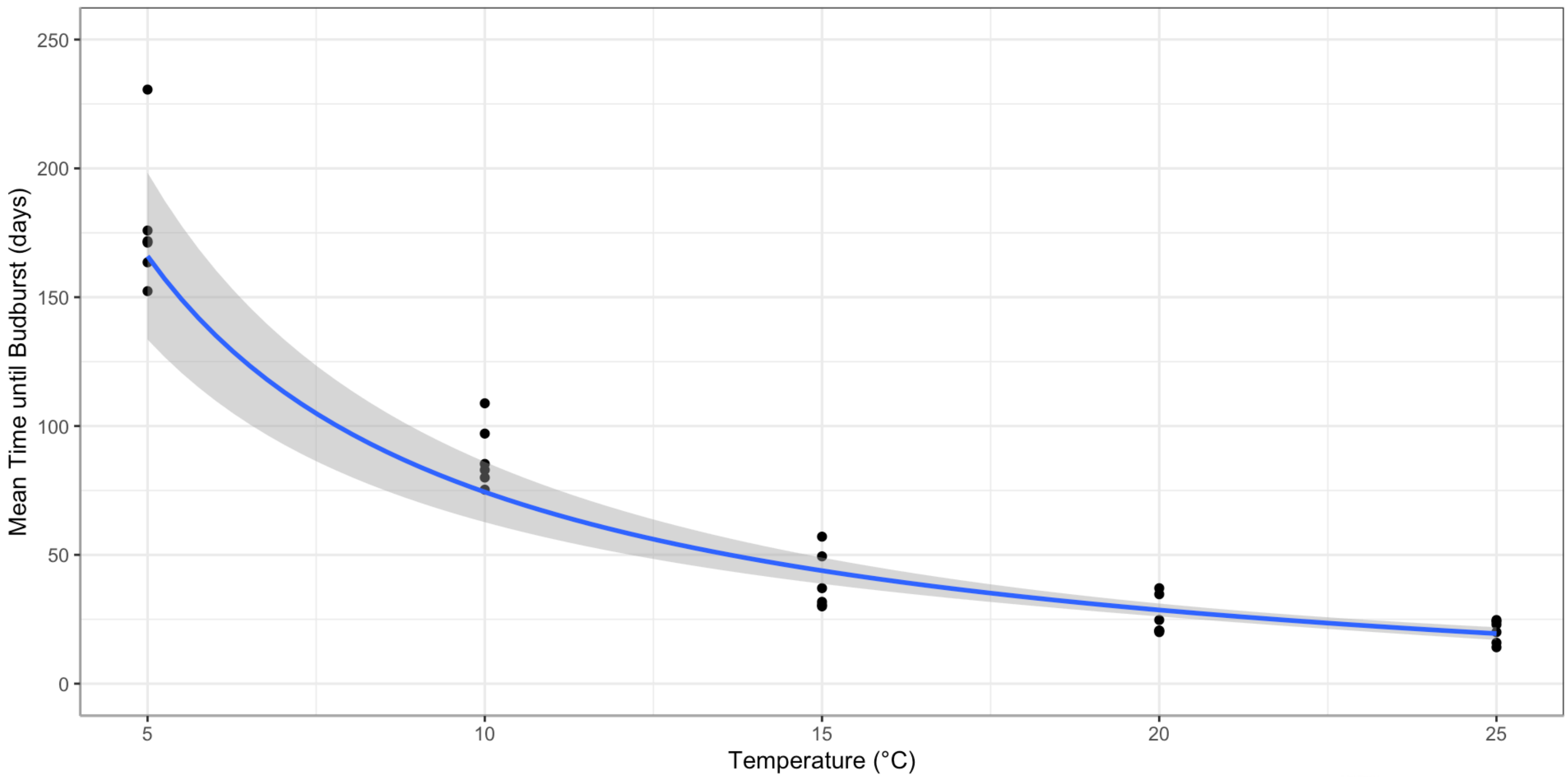
▶ expand R code



data from Charrier et al. (2011)

Fit using glm

▶ expand for R code

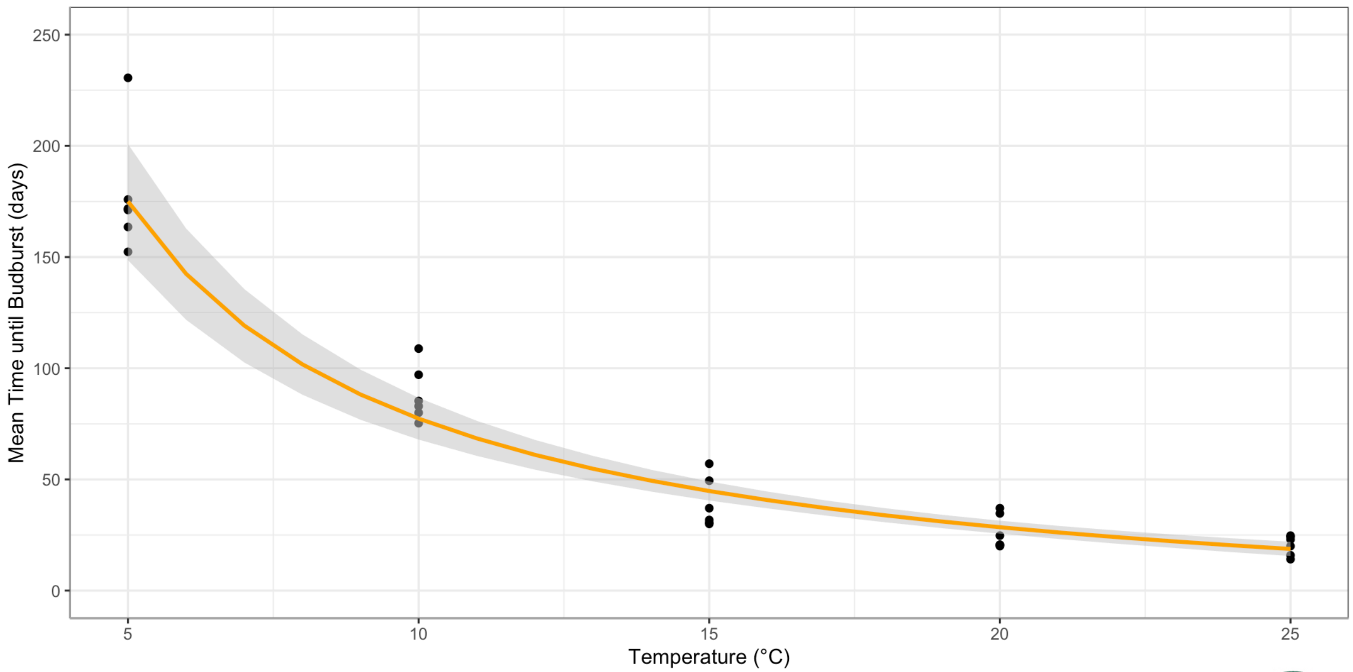


data from Charrier et al. (2011)

Fit using Stan

- expand for Stan code

- expand for R code



data from Charrier et al. (2011)

STATISTICS
George Mason University

Observational Data

STATISTICS
George Mason University

Observations (Marsham 1746-1958)

- Robert Marsham recorded the first occurrence of 27 signs of spring each year.
  - Marsham family continued to collect the data after Robert's death in 1797

- We model the day leaves first appeared on his oak trees.
  - spring.temp: cumulative daily temperature (2/1 to 4/30)
  - response.time: number of days from January 1 to budburst

- The first day of temperature accumulation, a, ma

STATISTICS
George Mason University

Observations (Marsham 1746-1958)

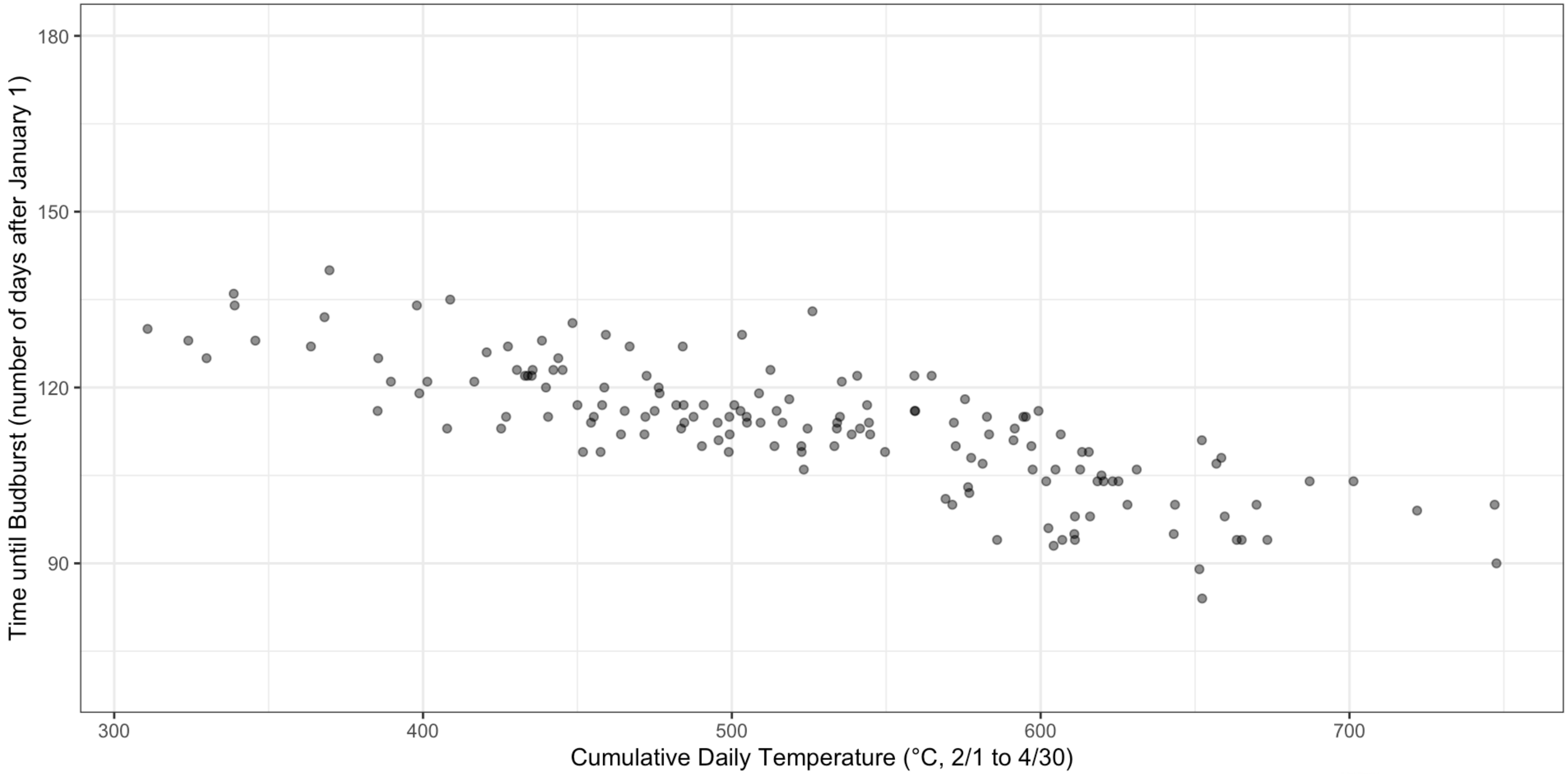
▶ expand for R code

year	response.time	spring.temp
1772	134	398.0
1773	113	541.5
1957	104	701.3
1958	122	472.4

data from Marsham (1746-1958)

Observations (Marsham 1746-1958)

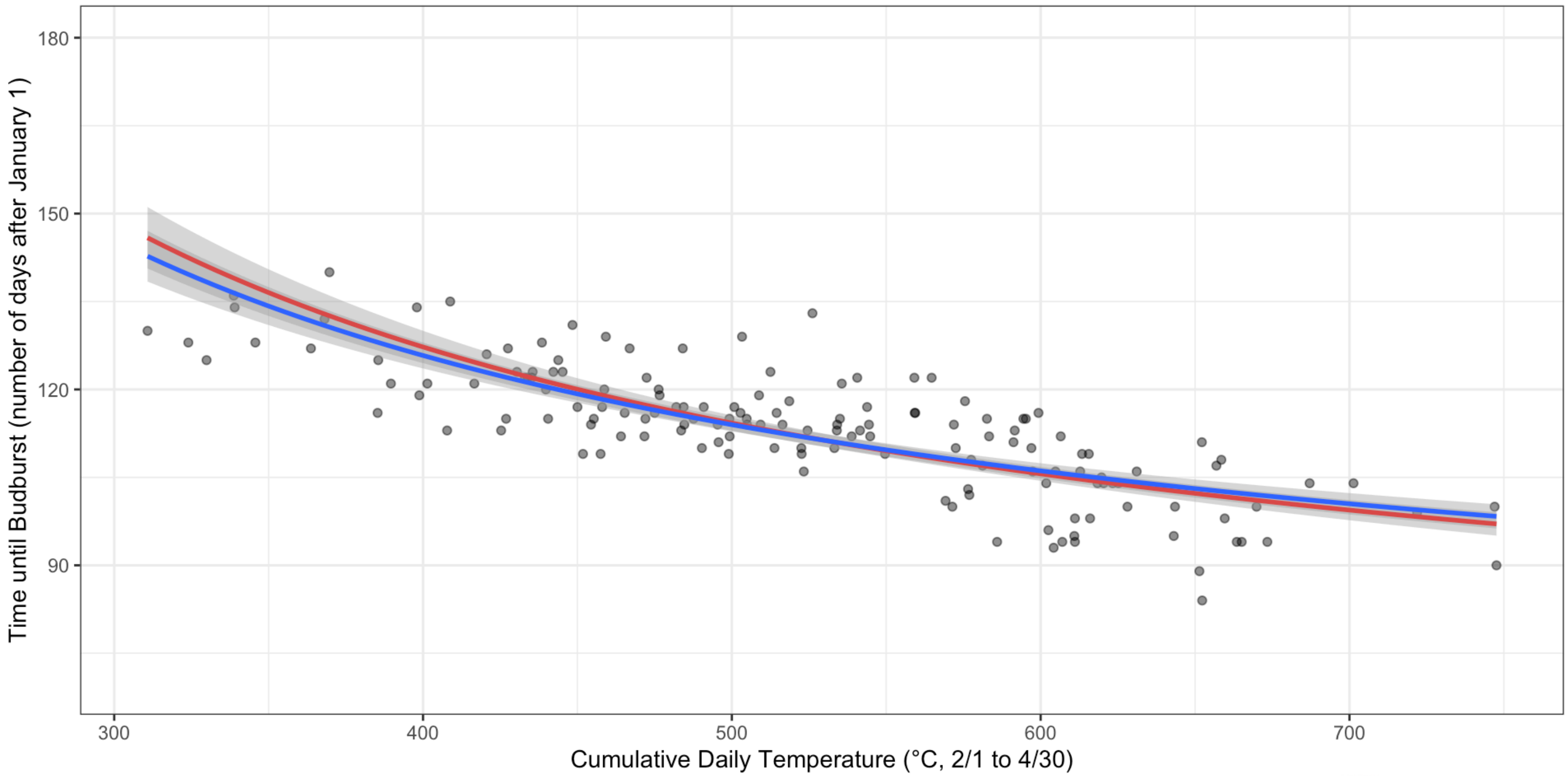
▶ expand for R code



data from Marsham (1746-1958)

Fit using lm/glm

- expand R code



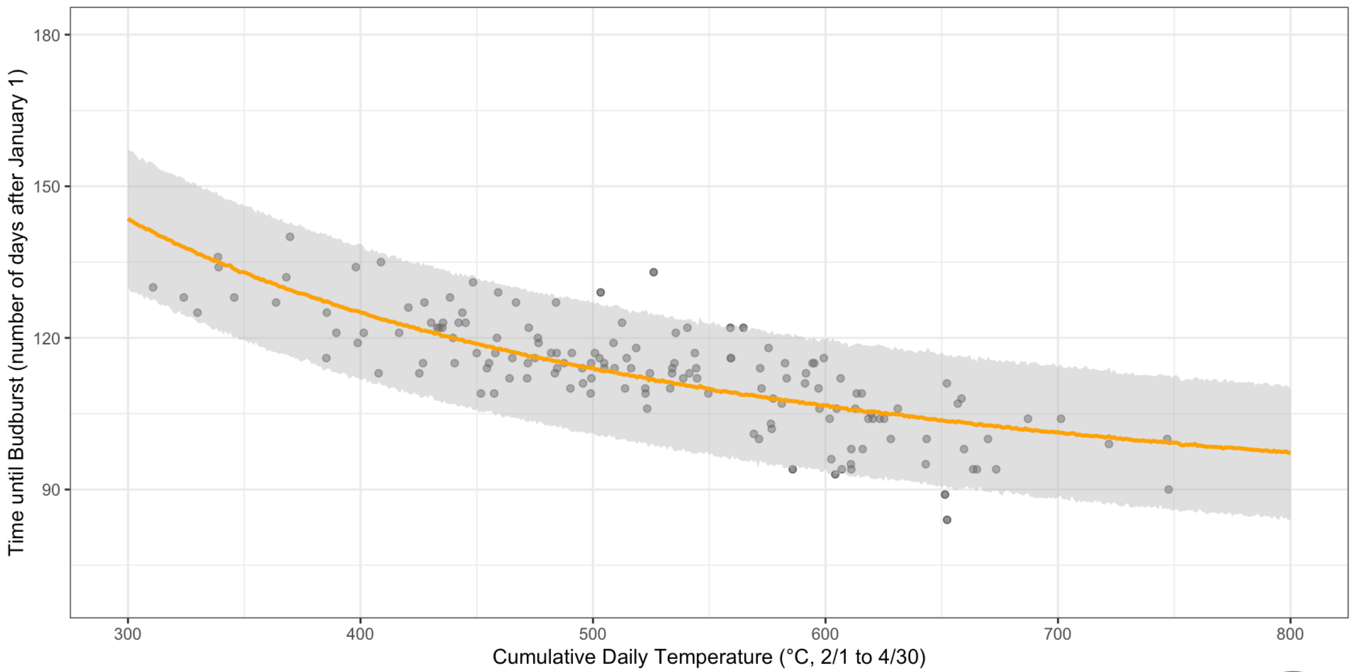
data from Marsham (1746-1958)

STATISTICS
George Mason University

Fit using Stan

▶ expand for Stan code

▶ expand for R code



data from Marsham (1746-1958)

STATISTICS
George Mason University

Non-asymptotic model

STATISTICS
George Mason University

One last takeaway from CLT

- In experimental setting, day forcing begins known by design.
  - For observational, day forcing begins not known.
  - Cumulative temperature from 2/1 to 4/30 is a proxy for total force. Proxy is accurate when γ large. Why?

- Recall $n_\gamma \dot{\sim}$ Normal $\left(\gamma/\mu + a,\ \gamma\sigma^2/\mu^3\right)$ when γ is large.
  - When γ is large, n_γ is large, and $n\mu \approx \sum_{i=a}^{n} \mu_i \approx \sum_{i=b}^{n} \mu_i$ for any $a, b << n$. i.e., First few μ_i don't matter.
  - Argument fails when γ is not large.

STATISTICS
George Mason University

Non-asymptotic model

- Assume instead $X_i > 0$ is normal with mean μ_i and variance σ^2.
- Note that $\mathbb{P}\left(n_\gamma \leq m\right) = \mathbb{P}\left(S_a^m > \gamma\right) = \Phi\left(\frac{\sum_{i=a}^m \mu_i - \gamma}{\sqrt{m-a}\,\sigma}\right)$ so that the likelihood contribution of each observation is

$$\mathcal{L}\left(\gamma, \sigma, a; n_\gamma, \{\mu_i\}\right) = \Phi\left(\frac{\sum_{i=a}^{n_\gamma} \mu_i - \gamma}{\sqrt{n_\gamma - a}\,\sigma}\right) - \Phi\left(\frac{\sum_{i=a}^{n_\gamma - 1} \mu_i - \gamma}{\sqrt{n_\gamma - 1 - a}\,\sigma\right.$$

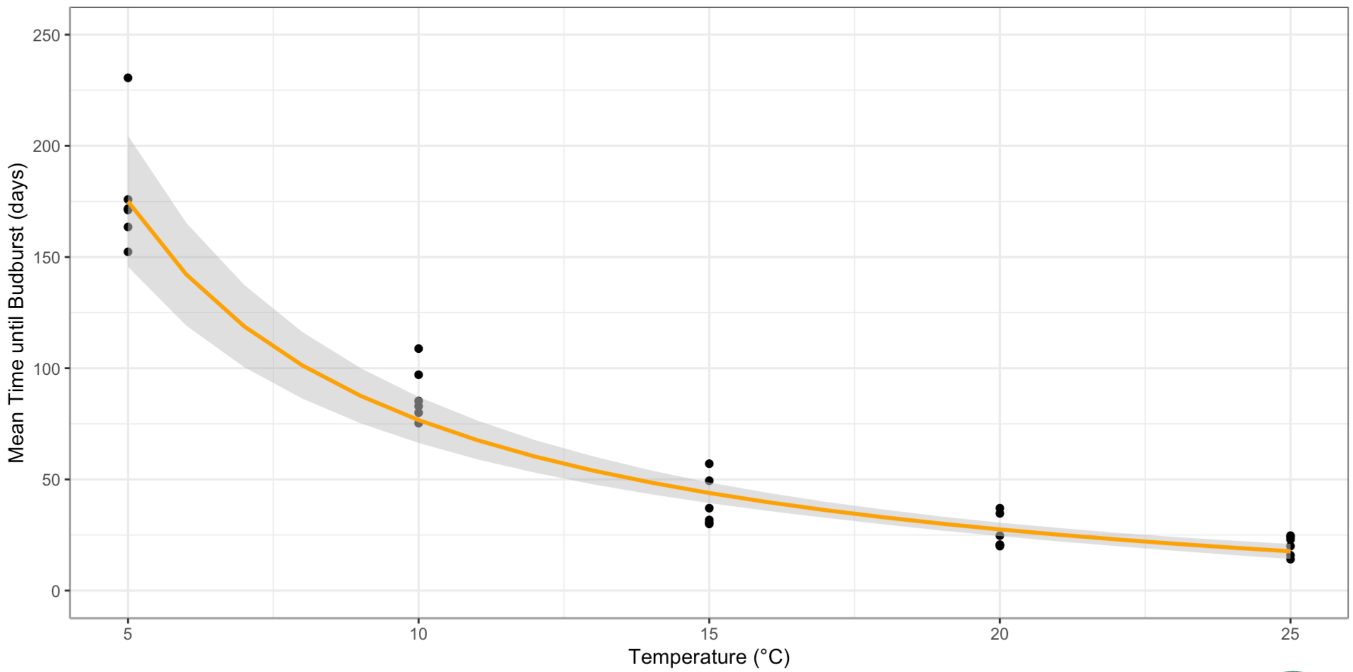
- n.b. $\sum_{i=a}^{n_\gamma} \mu_i = \sum_{i=1}^{n_\gamma} \mu_i \, 1_{i \geq a}$ and

$1_{i \geq a} \approx (1 + \exp(-b(i-a)))^{-1}$ when b is large.

STATISTICS
George Mason University

Experiment (Charrier et al. 2011)

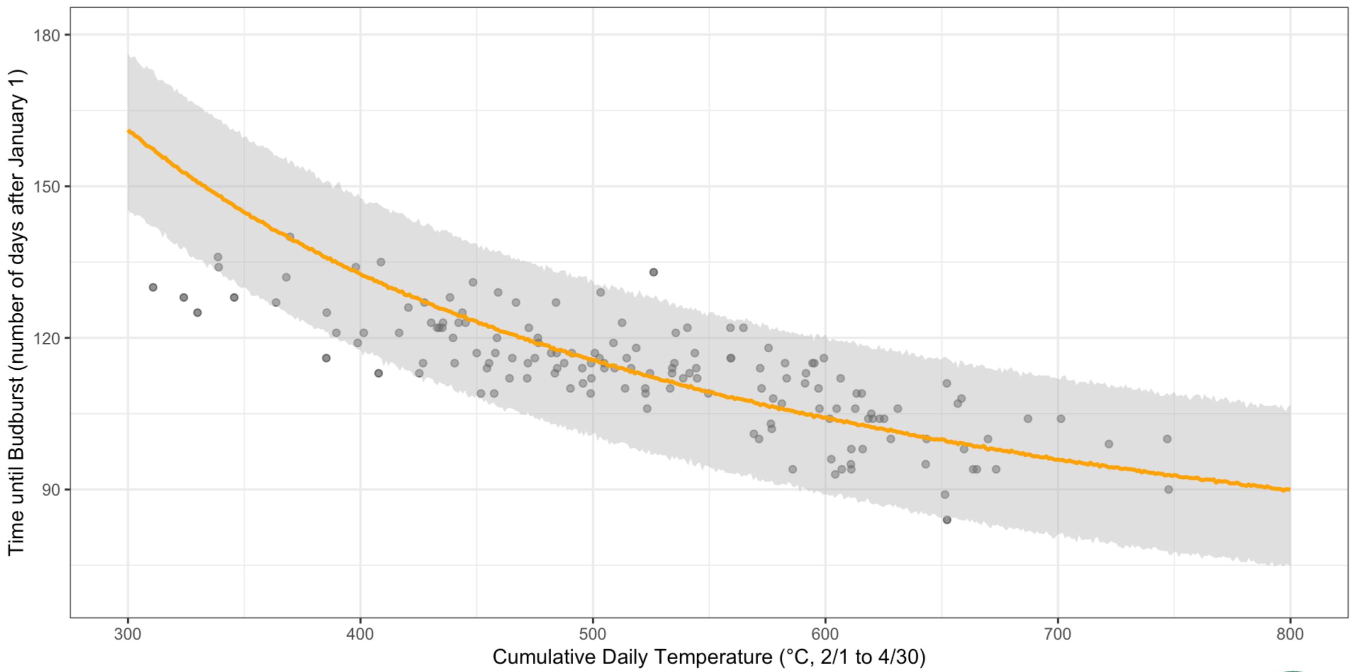
- ▶ expand for Stan code
- ▶ expand for R code (Experimental)



data from Charrier et al. (2011)

Observational (Marsham 1746-1958)

- expand for Stan code
- expand for R code (Observational)



data from Marsham (1746-1958)

STATISTICS
George Mason University

Conclusion

STATISTICS
George Mason University

Model comparison

- expand for R code

data	model	$\widehat{\gamma}$	$\widehat{\gamma}_{\text{lower}}$	$\widehat{\gamma}_{\text{upper}}$
Charrier	asymptotic	977	806	1144
Charrier	non-asymptotic	983	790	1175
Marsham	asymptotic	246	213	281
Marsham	non-asymptotic	380	344	415

STATISTICS
George Mason University

Summary

1. We justified modeling biological process as stopped random walks.
2. We reviewed the CLT for stopped random walks.
3. We applied the CLT to experimental and observational data.

- Found CLT approximation compared well to non-asymptotic model.
  - The model can be complicated to allow for more covariates or additional variation.

STATISTICS
George Mason University

References

1. Auerbach, Jonathan. (2023). A demonstration of the law of the flowering plants. Real World Data Science. https://realworlddatascience.net/ideas/tutorials/posts/2023/04/13/flowers.html

2. Charrier, G., Bonhomme, M., Lacointe, A., & Améglio, T. (2011). Are budburst dates, dormancy and cold acclimation in walnut trees (Juglans regia L.) under mainly genotypic or environmental control?. International journal of biometeorology, 55(6), 763-774. https://pubmed.ncbi.nlm.nih.gov/21805380/

3. Marsham, R. (1789). XIII. Indications of spring, observed by Robert Marsham, Esquire, FRS of Stratton in Norfolk. Latitude 52° 45'. Philosophical Transactions of the Royal Society of London, (79), 154-156.

STATISTICS
George Mason University