Thoughts on an improved temperature model

Lizzie, so far

September 8, 2020

Contents

1 What we aim to do

We want to improve our survival model with a better temperature model. As Mike explained (12 July 2020 email):

In general I'd say the big scientific input on the temperature side of things would be:

a) What temperature summary is the best input to the phenology model, i.e. T_{mean} or T_{min} or T_{max}, etc.
b) How do the temperature recordings relate to that summary, for example the recoding at any given time during the day could be considered a "noisy" measurement of that summary.
c) How accurate are the temperature recordings

Basically we need to know how much to smooth the temperature time series in order to convert the discrete readings into something continuous that we can input into the survival model.

2 Where I think we're at now ...

Mike wrote on 12 August 2020, "One thing to keep in mind is that there are many levels of detail with which we can model these effects. If we assume the relationship between recorded temperature and meaningful energy input are so complicated that its hard to build an explicit model then that differences can just be absorbed into the measurement variability. In other words wed do a separate spline fit for each plant to account for all of that heterogeneity. The next level of complexity would be to have varying parameters everywhere, and then at some point we could have a more mechanistic model. Even if we cant construct an explicit mechanistic model, however, reasoning about it conceptually should give us a wealth of information about time scales and other orders of magnitude from which we can construct principled prior models for the cruder models. Lots of exciting science to do!"

And on 4 September 2020 (in reply to me sending a version of this document):

"Milad's spline introduction (included as one of the links [from Ailene, see Splines in Stan]) is definitely one of the better introductions out there, let alone for applications in Stan.

One thing to keep in mind is the we can always separate some of the local temperature effects from the temperature model. In particular if we have a high-level temperature model that captures seasonality + trend + weather fronts then shading and other such effects can be modeled with varying intercepts between the "temperature" and what goes into the Wang-Engels model. They could also be added as intercepts between the Wang-Engels output and the survival probability; which is more appropriate depends on whether or not the effect is better modeled as modifying the input temperature or the input energy."

3 Plant temperature

Continuing our discussion of how to improve our temperature model, I did some reading in grapevine canopy and berry temperature, especially compared to the temperature data that we have—measured air temperature.

This is a well-studied topic. People have been using basic energy flux models for this since at least **?** and developing them here and there into models growers can use to estimate berry temperature (e.g., **?**). There are also continuing comparisons of how canopy and berry temperatures relate to measured air temperature (e.g., **??**). After all those years, though I am not sure there is an agreed-upon model (I found the **?** the best paper I read). Papers from Washington state diverged from those in Chile, and these's not even consistent info on 'our temperature station was x far away and in this type of location,' so I feel a little uncomfortable doing too much more than thinking on where and how to let the plant temperature vary from the air temperature.

That said, I think I gleaned a few insights useful to our priors and our temperature model.

- Sun and shade matter to berries: **?** found exposed fruit was 1.4-7.3 C higher than measured air temperature and shaded fruit were 0.5-4.4 C lower. (Relatedly, most radiative heat for berries goes into sensible, so you can ignore latent heat etc.)
- Sun and shade matter to leaves also: **?** found exposed leaves were similar to measured air temperature and shaded leaves were 0.9 C lower. (Relatedly, most latent heat matters a lot with leaves.) **?** found canopy seemed to increase diurnal temperature range (canopy minima 1.2 C less than air and maxima 2.0 higher) though I am not sure how consistent this is as other studies did not mention it.
- Aspect and row positioning, as well as trellising matter to temperature (**?**), though I didn't find a consistent model of this. But it makes sense.
- From conversations with collaborators—cover crop matters, so if you have a cover crop between rows that will reduce temperature.... at least while it's alive. So does irrigation. And irrigation x cover crop: if you have overhead irrigation it keeps your cover crop alive longer and keeps temperatures lower.
- Lags: I found conflicting info on whether plant temperature lags behind air temperature (or vice versa I guess). **?** found a lag of several hours for max temp comparing canopy (later max, around 17:00) and air (high from 11-17:00) while **?** reported no lag.

Here's how this could matter to our model:

- Several of these factors are vineyard-specific (trellising, irrigation, generally cover crop) and some are block-specific (aspect, row positioning), suggesting we might want to let the temperature model vary at these levels if possible.
- Some of these factors vary across the season: the cover crop often dies later in the season and growers often remove leaves to increase light/warmth in cool climates later in the season. But then, other places leave the leaves up .... not sure we want to deal with this.

– We often assume 40 C is the upper limit for most plant processes, as proteins denature above that, but we might want to think about where to set that prior (i.e., make sure it is on plant-temperature, not air temperature).
– I think we can assume no lag, since there is not a known one.

4 Lizzie's early-August thoughts

I think the sine curve is not so bad. If we want to improve we could consider a spline (or something similar) **to capture weather fronts,** which is our main issue (to me, at least, from looking at the residuals). I would also like to **add a baseline long-term trend** since most locations are getting warmer with climate change. Even if we don't see it here I'd be interested in having it in the model soon-ish as we would want it in most models.
Mike's reply (4 August 2020):

> The challenge is in accurately modeling the splines. People use splines/GAMs/etc as black boxes and take for granted that they work, but using them in a more principled way is much more challenging. See for example `https://rss.onlinelibrary.wiley.com/doi/full/10.1111/j.1467-9868.2011.00777.x` the kind of work needed to establish spine priors that accurate capture domain expertise about the possible interpolation functions. In practice we might be able to sneak around some of the troubles with heuristics but it will definitely take some work.
>
> What we'll really need to implement the splines properly is information about what the splines are trying to capture.
> - What are the time scales for weather fronts and other features that we want the splines to capture? Hours? Days? Weeks?
> - How much can those features influence the temperature? 1C? 5C? 10C?
> - How should we relate the recorded temperature to the local temperature around each plant? How much meaningful spatial and temporal variation is there between temperature measurements? I think a lot of this is taken for granted when people model Wang-Engels as contributions per day where all of this variation gets averaged out to a single contribution that correlates with the recorded temperature.

And he continued on 5 August 2020:

> ...how are the temperature measurements related to the energy input of the plants? Is there appreciable measurement error? Is there appreciable variation in where and when the measurements are made and the local environment around each plant? If the plant is integrating energy then how does that integral relate to the measurement at a single time during the day? Does that relationship change with season and geography?
>
> We can add fudge factors and the like but the more phenomenological of a story we have the better prior models we can impose. My hunch is that the actual phenolog-

ical transitions aren't super well identified and so any substantial variation in the temperatures will be degenerate with the energy accumulation model, but we'll have to bear that out.

One thing that might be worth trying – since we're moved to a survival approach the hazard function doesn't need to be as continuous as the in the threshold/Wang-Engels model, so it might be possible to replace the hazard integral (and the need to carefully interpolate) with a hazard sum without breaking the gradients. I'll have to play around with this once I'm settled in with my new machine and caught up on everything I've missed.

5 Overview of options considered

1. Fourier: I love Fourier, but I am not sure what we'll get from this beyond the sine curve for climate data here.
2. Autoregressive: Maybe, but I worry these will require a lot of decisions we're not sure of—the lag may vary over time, the timing of seasons varies ... they feel not flexible enough to capture weather fronts.
3. Sine curve: Seems pretty good as a first pass, but looks to miss weather fronts.
4. Splines: this is just a catch-all term for smoothing f(x)s. They seem to usually chop up the time-series to create some bends. I have seen a lot of climate folks use thin-plate splines, but I am still trying to figure out why.
5. GAMs: This is just a generalized additive model where the linear predictor uses some smoothing f(x)s.
6. ...

6 Emails and other notes

6.1 Christy Rollinson email (July 2020)

With the temperature modeling are you predicting past or future? I have a couple approaches I've used for gap filling and spaitio-temporally downscaling daily and subfamily meteorology data with uncertainty, but they're probably overkill. I've been using splines in GAMs a lot for this because of the nice tradeoff between being fairly computationally quick, ability to capture "surprises" and a variety of shapes among sites with the same function, and ability to robustly estimate uncertainty.

6.2 Ben Cook email (July 2020)

As for your query, I guess it depends on exactly what you want to to do. By "model the temperature data", do you mean you want to generate a new synthetic time series (say, daily data for one year) sampled from the existing data? One approach to do this would be to, using the multiple years of data you have, generate an average daily climatology, and. then convert each year to ANOMALIES. Once could then resample from the anomalies (e.g., using an autoregressive approach) and then superimpose these new synthetic anomalies back on to the climatology. You would probably want to fit a different autoregressive model to each season (DJF, MAM, JJA, SON).

6.3 From Ailene (July 2020)

Lizzie chatted with Ailene in late July. Ailene was very pro-splines and just sent some links.

Splines are the heart of the brms non-linear modeling (see here).

Link about GAMs (in brms).

Splines in Stan.

6.4 Random references from Lizzie

A Spline Model of Climate for the Western United States for monthly data, from 2006

For daily rainfall but seems to have a good review.

Another article with a semi-useful intro.

A 2019 thin-plate spline paper

6.5 From Iñaki email (July 2020)

The only paper I can propose to read is the chapter we wrote with Isabelle about phenology models (see attached file). There is an important review of different models used to simulate phenology stages but it was quite focused on dormancy (endo and eco) phases. There is also this one from Rebaudo et al. (focused on insects).

Usually we tested last years (with Amber and other colleagues): GDD, Sigmoid, Richardson, Wang, Normal (also called Chuine), triangle. There are other interesting as Logan (1, 2, 3), Anderson, Sine. But I did not tested them for plants and they are very similar to other models.

Lizzie notes: the two files are in the phenmodels/temperatre refs folder, but I am not sure they include temperature models.
References

Cola, G., O. Failla, and L. Mariani. 2009. Berrytone-a simulation model for the daily course of grape berry temperature. Agricultural and Forest Meteorology 149:1215–1228.

Costa, J. M., R. Egipto, A. Sanchez-Virosta, C. M. Lopes, and M. M. Chaves. 2019. Canopy and soil thermal patterns to support water and heat stress management in vineyards. Agricultural Water Management 216:484–496.

Millar, A. A. 1972. Thermal regime of grapevines. American Journal of Enology and Viticulture 23:173–176.

Pena Quinones, A. J., G. Hoogenboom, M. R. Salazar Gutierrez, C. Stockle, and M. Keller. 2020. Comparison of air temperature measured in a vineyard canopy and at a standard weather station. Plos One 15.