Overall notes on 22 March 2020

Current process model of leafout day (n) varies as a function of average daily temperature (X_i). In this model leafout occurs after accumulated daily temperatures cross a critical threshold (β). Ideally, researchers would know the date that temperatures generally start to accumulate, and accumulate from that zero point to n. In practice researches often accumulate over a fixed window $([a, b]$ such as March 1 to April 30) that they apply to all species or sites. Thus, let:

$i =$ index the days, $i = 0, 1, ..., N$

$X_i =$ observed temperature on day i, assume $X_0 = 0$

$\mu * i =$ average temperature on day $i = 1; X_i \sim \mathcal{N}(\mu * i, \sigma), i > 0$

$[a, b] =$ temporal window over which temperature is measured, where $0 <= a < n <= b <= N$

$$S_a^b = \sum_{i=a} b X_i$$

$\beta =$ a threshold of interest, $\beta > 0$, (for example, F* or required GDD)

$n =$ the first day such that $\beta < S_n$, (for example, day of year (doy) of budburst)

Here, X_i is a gaussian random walk with drift and n is a hitting time. Thus, the continuous time generalization would be a Brownian motion with instantaneous variance σ and drift μ. The time, t, the process hits some threshold β is distributed Inverse Gaussian Distribution, having mean μ/β and variance $\mu * \sigma/\beta^3$. Assuming $a = 0$ and $n = b$, then regressing n (e.g., day of leafout) against average daily temperature until n (as much research does when calculating temperature sensitivities currently), then this is equivalent to: $S_a^n/n \approx \beta/n$. Thus, $log(n) \approx log(\beta) - log(S_a^n/n)$ and $log(n)$ is linear in log-average daily temperature with a slope -1 and intercept $log(\beta)$.

Q&A:
Q: Does $X_i \sim \mathcal{N}(\mu * i, \sigma), i > 0$ mean that each day temperature increases by μ? (If yes, what if it is not so linear?)
A: Yes, μ is the slope or change in the daily average. If the daily average is not linear in i, it would be difficult to relate n to μ.

Q: Are we still happy with this hitting time model given $a = 0$ and $b = n$?
A: You can generalize the hitting time model to any interval $0 < a < n < b < N$. Although to make your point about a nonlinear relationship, I think it's easier/clearer to assume a = 0 and n = b. You can wave your hands/simulate for more complex models.

Q: How the hell do you jump to $b * (b + 1)/2 - n * (n + 1)/2$?
A: The sum of the integers from 1 to n can be written $n(n + 1)/2$. Legend says that Gauss derived the formula in elementary school when asked to sum the numbers from 0 to 100. When I write $S_a^n = \mu(n(n + 1)/2 - a(a + 1)/2)$, I'm summing $\mu * i$ from $i = a + 1$ to n.

Q: Given this formulation, would taking the log x and y (aka, n and mu) when plotting them against each other still linearize things?

A (extended): Sorry, I must have thouyght we wanted to derive the relationship between S_a^b and n. e.g. in the bottom of Figure S1 it looks you plot n against average daily temperature until n, $S_a^n/(n-a)$. Assuming for the moment that $a \approx 0$, $S_a^n/(n-a) \approx S_a^n/n \approx \beta/n$.

Thus, $log(n) \approx log(\beta) - log(S_a^n/n)$ and $log(n)$ is linear in log-average daily temperature with slope -1 and intercept $log(200) = 5.3$, which both appear to agree with your plot.

For μ and n, taking logs also approximately linearlizes things. For μ and n, you can write:

$\beta \approx \sum_{i=a}^{n} \mu * i = \mu \sum_{i=a}^{n} i = \mu(n(n+1)/2 - a(a+1)/2)$

Letting $\gamma_1 = 2 * \beta$, $\gamma_2 = -a * (a+1)$ and $\gamma_0 = \gamma_1 * \gamma_2$

$1/\mu = \gamma_0 + \gamma_1 * n * (n+1)$

Strictly speaking, $1/\mu$ is linear in $n * (n+1)$ and you should plot $1/\mu$ against $n * (n+1)$ BUT taking logs is probably fine since if n is large enough so that $n^2 \approx n * (n+1)$ and a is small enough so that $\gamma_2 \approx \gamma_0 \approx 0$, taking logs of both sides gives:

$-log(\mu) = log(\gamma_1) + 2 * log(n) = log(2 * \beta) + 2 * log(n)$

so $log(\mu)$ is roughly linear in $log(n)$ with slope -2 (or $log(n)$ is linear in $log(\mu)$ with slope -1/2).

From J. Auerbach on 9 March 2020 ...

In decsenssupp.pdf, X_i is the observed daily temperature, not the average. $\mu * i$ is the average temperature on day i so that μ is the slope or the change in the average daily temperature. Accumulated daily temperature from time a to b (where $0 <= a < n <= b <= N$) is denoted S_a^b. The critical threshold is β, which is crossed on day n when cumulative threshold S_a^n equals β. I think you can delete S_i and M_i in the documentI'm not sure what the notation $\sum_{X_1}^{X_n} X_i$ means–sorry if I introduced it.

[Rest of email was Q&A above.]

Original from J. Auerbach on 2 February 2020 ...

I figured the first step was to write n as a function of S_a^b. (Dividing by (b-a) shouldn't make a difference.) Then you can calculate variances, covariances, the regression slope, etc. (Although it's not obvious to me how to do this in closed form.)

It's actually easier to write S_a^b as a polynomial in n.

We defined n to be the day such that the cumulative temperature hit β:

$\beta \approx \sum_{i=a}^{n} X_i$

Assuming n is large, we can replace this sum with its average

$\sum_{i=a}^{n} X_i = \sum_{i=a}^{n} \mu * i$

So,

$\beta \approx \sum_{i=a}^{n} \mu * i = \mu * \sum_{i=a}^{n} *i$

and,

$S_a^b = \mu * \sum_{i=a}^{b} *i = \mu * \sum_{i=a}^{n} *i + \mu * \sum_{i=n}^{b} *i = \beta + \mu * \sum_{i=n}^{b} *i$

But $\sum_{i=n}^{b} *i$ is just the sum of the integers from n to b which is the sum of integers from 1 to b minus the sum of integers from 1 to n

$b * (b+1)/2 - n * (n+1)/2$

So,

$S_a^b = \beta + \mu * (b * (b+1)/2 - n * (n+1)/2)$

Letting $\alpha_1 = \beta + \mu * b * (b+1)/2$ and $\alpha_2 = \mu/2$

$S_a^b = \alpha_1 - \alpha_2 * n * (n+1)$

which is clearly not a straight line, and OLS (either S_a^b on n or n on S_a^b) is not a good idea.

Original from J. Auerbach on 22 January 2020 ...

Sorry, I'm confused by the additional notation. Is it correct to deonte the time window $[a, b]$ and $S_a^b = \sum_{i=a}^{b} X_i$? The question is then how a and b relate to n, the first time, i, that S_a^i hits β, i.e. $\sum_{i=a}^{n} X_i \approx \beta$.

I originally assumed $a = 0$ and $b = n$, in which case $M_b = M_n \approx \beta/n$.

If a, b are chosen independently of n, for example the entire year/spring, then the relationship would be similar:
Fix β so that:
$\beta \approx \sum_{i=a}^{n} X_i \approx \mu \sum_{i=a}^{n}$
or
$\mu = \beta / \sum_{i=a}^{n} = \beta/((n * (n+1) - a * (a+1))/2) \approx 2 * \beta/n * (n+1)$

Edited by Lizzie ...

Let:

$$i = \text{index the days}, \; i = 0, 1, ..., N$$

$$X_i = \text{temperature on day } i, \text{assume } X_0 = 0$$

$$\mu = \text{average temperature on day } i = 1; X_i \sim \mathcal{N}(\mu * i, \sigma), i > 0$$

$$S_i = \sum_{X_1}^{X_n} X_i, (\text{for example, GDD})$$

$$M_i = \text{cumulative mean}, S_i/i \text{and thus } M_n = S_n/n$$

$$\beta = \text{a threshold of interest}, \; \beta > 0, \; (\text{for example, F* or required GDD})$$

$$n = \text{the first day such that } \beta < S_n, \; (\text{for example, doy of budburst})$$

$$l = \text{the last date over which temperature is integrated}$$

'Sensitivity' is generally measured by taking the slope—I am calling it m here—of a linear regression of n against M_l.

It is not surprising that n decreases with μ (and thus M_l) under the null model (this is equivalent to saying it is not surprising that plants leafout earlier with warmer spring temperatures, we all are fine with this). Researchers are surprised that the decrease of m as μ increases and suggest this is evidence of declining sensitivities; n is "reacting" or "changing" to M_l.

So I say (I think, assuming the math still holds): Plotting n against M_l is similar to plotting n against S_n/n, which is approximately the same as plotting n against β/n, which is a hyperbola. The data is consistent with the null model; no need for n to react to M_l.

Original from J. Auerbach on 15 January 2020, with corrections from 22 Jan 2020

...

Let:

$$i = \text{index the days}, \ i = 0, 1, ..., N$$
$$X_i = \text{temperature on day } i, \text{ assume } X_0 = 0$$
$$\mu = \text{average temperature on day } i = 1; X_i \sim \mathcal{N}(\mu * i, \sigma), i > 0$$
$$S_i = \sum_{X_1}^{X_n} X_i$$
$$M_i = \text{cumulative mean}, , S_i/i \text{and thus } M_n = S_n/n$$
$$\beta = \text{a threshold of interest}, \ \beta > 0$$
$$n = \text{the first day such that } \beta < S_n$$

Note: a time n, we have $M_n = S_n/n \approx \beta/n$.

Some say: It is NOT surprising that n decreases in μ (and thus M_n) under the null model. It IS surprising that the decrease of n in M_n slows under the null model. This may be evidence of declining sensitivities; n is "reacting" or "'changing" to M_n.

You say: Plotting n against M_n is the same as plotting n against S_n/n, which is approximately the same as plotting n against β/n, which is a hyperbola. The data is consistent with the null model; no need for n to react to M_n.

If this is the correct framing of the problem, X_i is a gaussian random walk with drift and n is a hitting time. I believe the mean and variance are difficult to work out in closed form since you would have to integrate S_n over n. However, the continuous time generalization can be worked ot: a Brownian motion with instantaneous variance sigma and drift mu. The time, t, the process hits some threshold beta is distributed Inverse Gaussian Distribution, having mean μ/β and variance $\mu * \sigma/\beta^3$. This seems like a more accurate description of the problem anyway since the relationship between temperature and time is not discrete.