Let’s start from the discrete definition of the SR2009 model in their SM, section 2. Let’s ignore controlling variables and error terms throughout. The original formula has subscripts i, for location, and t, for year. We will also drop those since we’ll need to dissect time inside a year.

$$Y=\sum_{h=-5}^{49} g(h+0.5)[\Phi(h+1)-\Phi(h-1)] \quad \text{eq. 1}$$

Here, Y is the log yield and h means heat, a certain temperature level. They confusingly used T for something else in their definitions. g(h) is the impact function they are trying to estimate, which will be piecewise linear on h in our case, but more on that later. The function $\Phi_{it}(h)$ is “the cumulative distribution function of heat”. The other document I sent, SR2006, defines it a little better in the context of our eq 1: “Using Φ(h) as the cumulative distribution of time with temperatures less than or equal to h, the time temperatures fall in a 1deg heat interval (h, h + 1] becomes Φ(h + 1) − Φ(h)”.

So the term within brackets essentially means the total time temperatures fall within the 1deg interval starting in h. The g(h+0.5) is merely centering the impact function in the middle of that temperature interval. We will need to define Φ a little more in depth and will be carrying the h argument around a lot, so let’s just assume their intervals were centered on the actual h instead. This is of course not exactly of what they did but we are shifting everything by a mere half degree. Our working equation will then look like:

$$Y=\sum_{h=-5}^{49} g(h)[\Phi(h+0.5)-\Phi(h-0.5)] \quad \text{eq. 2}$$

Now, let’s define the term in brackets in a more tractable way, retaining the meaning of being the “time temperatures fall in a 1deg heat interval”, now centered on h. This will be easier if we consider we are evaluating this on a discrete interval of time dt, in days. The value of dt we are actually using is 0.01 days. So, at a certain time t, the time temperatures fall in an interval centered around h will depend of the temperature variation with time T(t) will be:

$$\theta_t(h)=\begin{cases} dt & if\,[h-0.5]<T(t)<[h+0.5] \\ 0 & otherwise \end{cases} \quad \text{eq. 3}$$

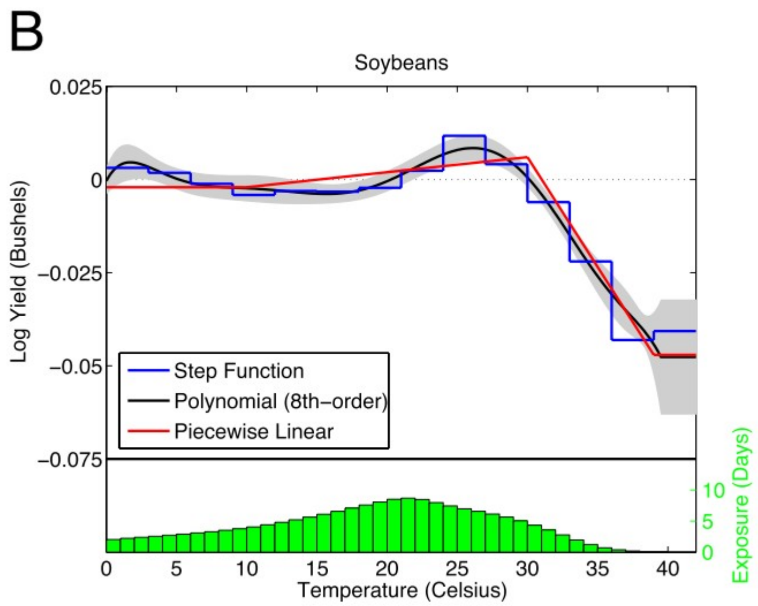
The total time temperatures fall in a 1deg heat interval centered in h will be the sum of θ over all time steps in a certain growing season GS:

$$\Theta(h)=\sum_{t\in GS} \theta_t(h) \quad \text{eq. 4}$$

So, our eq. 2 becomes:

$$Y=\sum_{h=-5}^{49} g(h)\Theta(h) \quad \text{eq. 5}$$

Now, let’s look into the piecewise specification of g(h).



We will focus here on the case where the effect changes linearly above a certain threshold K, which we will show later that can be related to EDD. This means we are dealing with the latter, descending, part of the red line on the figure after ~30 degrees, and assuming it's slope continues ad infinitum instead of becoming flat at around 40 degrees like the one in the figure. Temperatures in that range are very rare, so this should be an OK assumption. A more rigorous treatment should be given to the case where there is an upper bound since we also use when we are applying a GDD sensitivity instead of EDD, but let's leave it to another moment.

So, in the piecewise linear part of the function above a threshold K, g(h) is a simple slope and intercept funcion of h. Note that this is only evaluated for h above K:

$$Y=\sum_{h\geq K}(\beta_0+\beta_1 h)\Theta(h) \quad \text{Eq. 6}$$

We could in theory get β_1 and β_0 from digitizing the figure. However, the figure caption explicitly states that "Curves are centered so that the exposure-weighted impact is zero". This means, at least, that we can't trust a β_0 from this figure.

Here we can make the somewhat strong assumption that the impacts begin only after the threshold K, and therefore Y(K) = 0. This is not the case in the figure, and is even more problematic since we are dealing with the pieces in both sides of K which would mean assuming this is the left piece means the assuming it in the right piece is wrong. We have to check later what bias is this introducing.

But let's assume Y(K) = 0 for now. Then we can solve the linear function to get β_0:

$$\beta_0+\beta_1 K=Y(K)=0$$

$$\beta_0=-\beta_1 K \quad \text{Eq. 7}$$

Then the impact equation becomes:

$$Y=\sum_{h\geq K}\beta_1(h-K)\Theta(h) \quad \text{Eq. 8}$$

Now let's take a step back and consider the definition of degree days. Let's use just the one we are using for extreme degree days, which has only a lower but no upper threshold. Degree days accumulate only above a certain threshold K. A day with a temperature h one degree above K counts as one degree day. A day with h two degrees above K counts as two degree days, and so does two days with h only one degree above K. The concept can be generalized for an interval of time dt as:

$$DD_t = \begin{cases} 0 \quad if\ h < K \\ (h-k)\,dt \quad if\ K \geq dt \end{cases} \quad \text{eq. 9}$$

We can also define it discretized in h space using our definition of exposure in Eq. 3. DD for a interval dt is the sum of θ(h) for all h above K:

$$DD_t = \sum_{h \geq K} (h-K)\theta(h) \quad \text{eq. 10}$$

The total degree days above a threshold in the growing season, which we are calling EDD, is then the sum of DD for all the timesteps in the growing season:

$$EDD = \sum_{h \geq K} \sum_{t \in GS} (h-K)\theta(h) \quad \text{eq. 11}$$

Or, considering the definition of Θ(h) in eq. 4:

$$EDD = \sum_{h \geq K} (h-K)\Theta(h) \quad \text{eq. 11}$$

Substituting in the linear impact equation (eq. 8), we can find the impact of EDD in log yields simply as:

$$Y \ = \ \sum_{h \geq K} \beta_1 (h-K)\Theta(h) \ = \ \beta_1 \sum_{h \geq K} (h-K)\Theta(h) \ = \ \beta_1 EDD \quad \text{Eq. 12}$$