**Steps to Making a Phenology Model**

1. **Do a survey of the literature looking for laboratory temperature-developmental data (and other models)**.
   1. Summarize the mean developmental time and stdev for each stage of the insect. In some cases, you’ll just have data for each instar within a stage, others it will be summarized like “egg-adult” or “larval” or “nymphal” stages.
   2. Put all the data into an excel spreadsheet or a programmable stats program.
2. **Analyze the lab data on basic life history**
   1. First calculate the lower threshold and duration for each stage. Do this by plotting the mean developmental rate (mean developmental time)-1 at each temperature for each study.
      1. Expect with a variety of sources that there will be some variation in the developmental rate, even at the same temperatures. See if there is some sort of explanation for it in the methods – look for how often the data were checked, whether the temperature ranges were impossibly precise, etc.
      2. If there is a good reason to suspect that the data from a particular source are suspect, drop that data (if you consistently find deviations from 2-3 other studies, it suggests problems calibrating the growth chambers, methods used, etc.).



**Fig. 1.** Typical temperature development curve for insects. Generally have a linear portion, then a rapid drop off at higher temperatures.

* + 1. In the plot of the data, you expect that the developmental rate is linearly related to the temperature at which the experiment was run. However, at high temperatures, you expect to see a drop off of the developmental rate related to enzymatic deactivation (Fig. 1). You may see a slight departure at temperatures really close to the lower threshold for development as well, but this is of minor importance because it is overshadowed quickly as temperatures rise (it is also potentially a problem because if you run the developmental rate at a temperature very close to the lower threshold, you will end up taking a very long time to do the study and the checking of the developmental rate will likely not be every day (except in really old studies), so the error in developmental time is also much greater).
  1. Run a regression on the developmental time vs temperatures – ELIMINATING the non-linear portion of the curve. If you include all the high temperature data, you will skew the regression (make the slope flatter) and not describe the linear portion of the curve well (which is needed for the model) See. Fig. 1.
     1. The slope-1 of the regression gives the average duration of the stage in DD. In our example (Fig. 1), our slope is 0.0067253, so the duration is 148.7 DD (see the appendices).

**Fig. 1.** Regression of developmental rate versus temperature. A. Using all the data, the curve is flatter and doesn’t capture the drop off at the higher or the lower temperatures. B. the regression is done only with the temperatures below 30°C when the developmental rate began to drop.



* + 1. The lower threshold (LT) is = - intercept/slope. In our example, the intercept is (-0.025437) so the lower threshold is =0.025437/.0067253 = 3.78°C. (obviously, you can & should round some these final values).
    2. Typically, you can set the lower threshold using a single stage – stages that are not dependent on nutrition (e.g., egg stages, pupal stages) are better.
    3. While some people fit the regressions to each stage and say there is different thresholds for each, there is good evidence that is untrue (Jarošík et al., 2002; Jarošík et al., 2004)– just fit one and calculate the time period as in “c”.
  1. Go back to your summary of the different developmental times at each and make the times into DD for both the mean and the stdev.
     1. For example, if you know that the developmental time at 10°C was 22.77 days, you calculate the mean on a DD scale as (10-LT)\*22.77 or in our example as 141.6 DD.
     2. The standard deviation on a calendar date can be converted to a DD value for a particular temp data point by multiplying it by the same value (Temp-LT).
     3. If you have repetition at a particular temperature value, then you can approach this in a couple of ways.
        1. You can estimate the SEM based on the average DD for all the different temperatures (in the linear portion of the curve). In our case, if we do this, the mean DD for the egg stage is 149.7 DD with the SEM being 14.12575 (doing this, the standard deviation is the SEM because you are not estimating the deviations of every observation, but on the different mean observations at each temperature). The stdev would then be 14.12575 \* (13 in our case) or 50.93.
        2. If the developmental data actually has the stdev for each mean developmental time, after converting it to a DD basis, you can just take the average of the stdev – this might be a bit iffy if the developmental time is skewed for some of the temperatures, but it should be close.
        3. If you have the mean and stdev for developmental times, and the author actually indicates that they are normally distributed (or distributed by some other distribution), you can simulate the entire data set (even if not, you can do this and compare it to #2). So for example, if the data are normally distributed, and you have a mean DD and Stdev for each temperature, just use a random number generator for the normal distribution to estimate the sample size you want with a given mean and stdev. You can repeat this for each of the data points, and calculate the mean and stdev. Repeat this about 500-1000 times and you’ll have a bootstrap estimate of the mean and stdev.
     4. The importance of having the mean and stdev for a particular stage is that in situations where there are multiple stages and multiple generations, you can use variance addition to estimate when the different stages overlap and how much. The idea is that even though the egg stage may have a mean duration of 150 DD, there is a considerable variation around that value when any given individual may finish that stage. So you can’t just estimate that everyone in a cohort is done at 150 DD, they will emerge in the range of 150 ± *t*\*sd. If you are looking at the next stage, and it has a mean time to complete 250 DD and its own sd, then the variation for that stage is expected to be the sum of the variance (not stdev) of the first stage + variance of the second stage. The mean developmental period in the field would then be 150+250=400 DD, but the variability you would see in emergence time would be 400 ± *t*\*.
        1. I’ve used this approach on several models and it works pretty well. Especially useful if you have lab data for stages you don’t have in your field data. For example, you might have great data on the adult stage of a moth, but not good data for the larval stages. So you expect the emergence of the adult stage of the moth is defined by the logic above for the larval stages – this worked really well for WTLM, where the observed distribution of trap catches per generation was virtually identical to the lab predictions from developmental rates of the eggs, early and late instarts, and pupal developmental rates. I couldn’t estimate the first couple of generations from our field data, because the numbers were so low that they were hardly detectable using our sampling regime.
        2. This also works for things like aphids at least for a few generations. We had field data for RAA and AGA which have only a few generations per year on apples. By doing the variance added method, we could show that the distribution of aphid nymphs and adults caught over the entire season was well predicted when you considered the population growth per generation.
        3. Don’t expect this to work for >3-4 generations though! The overlap may become so severe that different generations would be indistinguishable – it’s a good starting point, but just may fail.
  2. **Estimation of the upper threshold.** There is a great paper on this (Dixon et al., 2009). They did a survey of the literature and found that the upper threshold across a wide range of taxa of pokilothermic animals and plants was 19.8°C higher than the lower threshold. If there is no lab data that gives a clue, this is a good standard to use, you’ll at least be in the ball park and if it appears to be way higher, then re-examine your data again to see if the lower threshold is reasonable.
     1. If you have temperature developmental data over a wide range of temperatures, you can look for the spot at which development starts to slow and use that value. Doesn’t have to be perfect, but just close, because the drop off of development at high temperatures is really quick – essentially if you inactive the rate limiting enzyme thermally, then that kills the animal. A range of just a few degrees is huge.
  3. **The type of upper threshold** – this is not a precise science! Everything should be a vertical cutoff (theoretically) – that is, if the temperature goes above a certain point, the chemical reactions that power development should stop. However, often times the horizontal cutoff (heat accumulations plateau at the upper threshold) works better at predicting phenology in the field. This is probably because the animals can thermoregulate a little behaviorally by moving to cooler areas (*e.g*., center part of the canopy or ground cover) when air temperature is high. You just have to try both types of thresholds and see how it fits the field data.
  4. **Determine the number of generations**. For some species this is incredibly easy, others not. A very simple way is to estimate the developmental time from the laboratory data as above as well as the variation possible. Then just use the seasonal DD accumulations to estimate how many generations are possible.
     1. You really need the phenology models to track generations, otherwise you may have 2 generations in a cold year and 3-4 in a hot year. Trying to just fit curves to the data gives really bad predictions when you just fit cumulative over the season because each generation generally increases the numbers caught dramatically.
     2. There are some situations where the number of generations per season are constant. Single generation/year or situations where the number of generations per year is fixed by photoperiod (e.g., Rosy apple aphid or apple grain aphid which have only 2-3 generations before they move to summer hosts (this is mediated by photoperiod).
     3. In some cases, there are also clear separations between generations (WTLM, OBLR, PLR, etc.). You can generally see that when you start to plot data out (especially in section 3di below).

1. **Field Data**
   1. **Properly designed phenology models:**
      1. Repetition across multiple locations that have different seasonal temperature profiles as well as repetition across years.
      2. Data taken frequently enough that each stage of interest will be sampled multiple times per generation. Typically, this is multiple times a week; once a week is probably a minimum for most pests at least in tree fruit. Things like aphids and mites need 2-3 times a week.
      3. Spray records should be available in most situations. If the pest is being looked at in a highly sprayed ecosystem (*e.g.*, pear psylla on pears), then you need multiple sites per year sampled frequently with spray records.
   2. **Weather data:**
      1. Always need accurate weather data. Data from stations 2-3 miles away may not be sufficient especially if the terrain is hilly and/or has lakes or large rivers nearby.
      2. Use within orchard loggers if possible. WITH SHIELDED temperature sensors!
      3. Use high quality gridded weather data (Daymet, other sources) if within orchard data is not available.
      4. Last resort to use AWN if station is >2-3 miles away and terrain is level. Make sure you have checked the 15 minute data, AWN doesn’t have records if there is missing data and this can cause problems when calculating the max-min temperature data. They also have frequent sensor errors.
   3. **Calculate DD:**
      1. Start with the thresholds from the literature data summary or the model if there is one already published
      2. I would normally start with looking at both a horizontal cutoff and a vertical cutoff and see which one looks better and has the least error associated with it.
      3. Initially, I would just look for cumulative DD since 1 January for each sampling date. If that doesn’t work, consider using a biofix, but have enough data that you can be sure the biofix actually improves predictions significantly.
   4. **Analyze the field data**
      1. *Initially, I would look at a weighted histogram.* That is for each date and location, merge the data for the cumulative DD with the number of individuals in each stage that you are monitoring. You would plot the cumulative DD and weight it by the numbers found in each stage being monitored.
         1. Look for breaks in the data that would correspond to different generations or overlap or places where you need more data. Experiment with looking at the bin widths, but don’t go too wide or you’ll lose the shape of the generations.
         2. It might not hurt to round your DD accumulations to the closest 5 DD so you get less noise – it doesn’t really change the fitting that you’ll do later on for the actual model.
      2. *Once you have the generations separated, start the process of fitting the data.*
         1. One thing you can do very simply is just look at box plots of the DD at which the different stages were collected (make sure to weight it by the number of individuals collected at each DD accumulation). This gives you an empirical distribution and will show skew and any extreme data points. You can use the standard box plots or for this purpose, it’s even better to use a box plot where you can set the whiskers to be some specific values (e.g., 5 or 95%) so that you can visualize and come back to this as you start making your model based on a specific statistical distribution. The box plot might also be an appropriate end point, depending on what you’re interested in and its super simple and informative.
         2. After the box plots, I typically start with a *p-p* probability plot (plots the observed probability density function versus the theoretical one), a *q-q* probability plot (plots the observed quantiles against the theoretical one (the *p-p* and *q-q* plots should be highly linear if the candidate distribution is a good fit).
         3. Plot the observed probability versus DD and the predicted probability for the fit distribution versus DD on the same axis – they should fit pretty well.
         4. I probably look at too many distributions as potential models. The majority of stuff fits the gamma, Weibull, or Gumbel distributions, with some fitting the normal distribution well (they would also fit the gamma very well in this case).
            1. Make sure you use maximum likelihood estimation.
            2. R probably has a bunch of options for this, Stata has several that I use that does all the analysis at once for each distribution that I choose and I can then decide between them.
            3. Stick to a single distribution between generations and stages – if nothing else, just for simplicity.
         5. Fit the data to only a portion of the full data set, then use the rest to evaluate how well the model predicts the part of the data set not used to generate the model.
            1. Multiple ways to evaluate this. My current thought is to evaluate the error in the model prediction and the observed data in the validation data set. For example, you might look at every data point in the observed validation data set and look for the average error on a DD basis (should show the average bias), look at the mean absolute deviation on a DD basis (shows the model repeatability).

You can also do this on a calendar date basis, but the kicker is that the error is not linear. An error of 1 day might be 5 DD near the start of the season, but 30 DD during the hottest part of the year and hotter locations. Basically, it will make you crazy, especially in the early season where you might have the same DD accumulation for several days in a row, then get one warm day and accumulate 5-10 DD. If you are looking for the time when you’d get to 15 DD, but the four days are below the lower threshold, you could be 5 days off, even though you were at 14 DD 4 day in a row.

* + - * 1. The “indifference band validation” was proposed by Steve Welch (Welch et al., 1981) where you plot two parallel lines that are offset from the line of observed=predicted. The lines were split by some difference that made no difference to pest management decisions - so like ±3 days or 5 days early and 3 days late depending on the pest. Welch plotted the observed date something happened (like 5% emergence or 50% emergence, etc.) versus when it was predicted to have occurred; to find the failure rate, just count the number of points outside the “indifference” bands compared to the total number of points and you have your error rate. You can also look at if there are points where error tends to accumulate to see if there are any systematic issues with each generation, or stage. Unfortunately, it has the problems associated with calendar dates mentioned above.

You could do the same analysis but on a degree-day basis, but in section 4a, that does it just as well, particularly if show the data using the plots I mention in the section dii.

* + - * 1. Another way to do this is to look at the probability of emergence from the model versus when it was observed (essentially the p-p plot). The advantage of this is: (1) you’ve already done it above, (2) you could modify the indifference band validation method to show how much error you’re interested in from a management perspective (e.g., you don’t care about a difference of 2-3% for egg hatch, but do care about >6%), (3) it can be used to show the difference based on the average difference and the mean absolute deviation as in 5a, and (4) it is more likely to directly indicate the thing of interest (e.g., reflects what the difference is in terms of egg hatch compare to DD or calendar dates which are not linearly related to the emergence curve or as heavily influenced by seasonal profiles of heat accumulation. I have not used this to date for a publication, but have been kicking it around for the pluses and minuses and think it should be considered as it gets rid of a bunch of concerns in a & b above.

1. Final check
   1. Plot the data on the same axis for the different stages and generations and see if it makes sense. Does it predict the box plots pretty well?
   2. Double check everything you did, especially evaluating the difference in errors between the model development and model validation data sets.
      1. Re-randomize the data and generate the model again and evaluate how different it is and how much the model validation changes.
      2. If the model based on the subset of data works, you can probably just use all the data for the final model. If you get a situation without enough data to fit the model to each instar, you can also lump instars together if you have a situation where young larvae versus old larvae are important from a management perspective.

**References Cited**

Dixon, A.F.G., Honěk, A., Keil, P., Kotela, M.A.A., Šizling, A.L., Jarošík, V., 2009. Relationship between the minimum and maximum temperture thresholds for development in insects. Functional Ecology 23**,** 257-264.

Jarošík, V., Honěk, A., Dixon, A.F.G., 2002. Developmental rate isomorphy in insects and mites. The American Naturalist 160**,** 497-510.

Jarošík, V., Kratochvil, L., Honěk, A., Dixon, A.F.G., 2004. A general rule for the dependence of developmental rate on temperature in ectothermic animals. Proc. R. Soc. Lond. B 271**,** S219-S221.

Welch, S.M., Croft, B.A., Michels, M.F., 1981. Validation of pest management models. Environ. Entomol. 10**,** 425-432.

**Appendices**:

**Example data in section 2.**

|  |  |  |  |
| --- | --- | --- | --- |
| **temp** | **egg** | **rate** | **dd** |
| 10 | 22.77 | .04391743 | 141.6 |
| 10 | 25.67 | .03895598 | 159.7 |
| 15 | 13.3 | .07518797 | 149.2 |
| 15 | 11.99 | .08340284 | 134.5 |
| 15 | 14.33 | .06978367 | 160.8 |
| 15.6 | 10.9 | .09174312 | 128.8 |
| 20 | 9.32 | .10729614 | 151.2 |
| 20 | 11 | .09090909 | 178.4 |
| 21.1 | 8.7 | .11494253 | 150.7 |
| 25 | 6.67 | .14992504 | 141.5 |
| 25 | 7.82 | .12787724 | 165.9 |
| 26.7 | 5.8 | .17241379 | 132.9 |
| 29.4 | 5.9 | .16949152 | 151.2 |
| 30 | 6.6 | .15151515 | 173.1 |
| 30 | 5.58 | .17921147 | 146.3 |
| 32.2 | 6.6 | .15151515 | 187.6 |
| 33.3 | 6.17 | .16207455 | 182.1 |
| 35 | 7.75 | .12903226 | 242.0 |

**Regression analysis in section 2.**

**Regression of rate vs temp not censoring the temperatures >=30°C**

. regress rate temp

Source | SS df MS Number of obs = 18

-------------+---------------------------------- F(1, 16) = 71.02

Model | .027269036 1 .027269036 Prob > F = 0.0000

Residual | .006143537 16 .000383971 R-squared = 0.8161

-------------+---------------------------------- Adj R-squared = 0.8046

Total | .033412573 17 .001965445 Root MSE = .0196

------------------------------------------------------------------------------

rate | Coef. Std. Err. t P>|t| [95% Conf. Interval]

-------------+----------------------------------------------------------------

temp | .0049791 .0005908 8.43 0.000 .0037266 .0062316

\_cons | .0042353 .0141755 0.30 0.769 -.0258155 .0342861

------------------------------------------------------------------------------

**. regress rate temp if temp<30**

Source | SS df MS Number of obs = 13

-------------+---------------------------------- F(1, 11) = 192.53

Model | .021138972 1 .021138972 Prob > F = 0.0000

Residual | .001207752 11 .000109796 R-squared = 0.9460

-------------+---------------------------------- Adj R-squared = 0.9410

Total | .022346724 12 .001862227 Root MSE = .01048

------------------------------------------------------------------------------

rate | Coef. Std. Err. t P>|t| [95% Conf. Interval]

-------------+----------------------------------------------------------------

temp | .0067253 .0004847 13.88 0.000 .0056585 .0077921

\_cons | -.025437 .0096852 -2.63 0.024 -.046754 -.0041201

------------------------------------------------------------------------------

. summarize dd if temp<30

Variable | Obs Mean Std. Dev. Min Max

-------------+---------------------------------------------------------

dd | 13 149.7321 14.12575 128.838 178.42