Phenology project model description

Louie Yang
Collin Edwards

October 31, 2016

1 Overview

2 Predictions

- Year-to-year variation – (early+late) vs normal phenologies – if day-to-day variation is low, annual variation is informative about phenology, so cumulative temp and temp will be favored, while photoperiod will be disfavored. Cumulative precip may be informative of seasonal patterns, while instantaneous precip will be disfavored.

- Day-to-day variation – smooth vs rough variation – with very noisy variation, photoperiod and cumulative temp will be favored as a long-term approximation of season trends and a way to avoid false starts. Instantaneous temps and precip will be disfavored as too noisy.

- Spatial variation – Depends on the balance of inter-annual and intra-annual variation in the locations being compared. General prediction is that day-to-day variation creates seasonal noise, and will favor photoperiod and cumulative measures and disfavor instantaneous measures, while year-to-year variation is informative about phenological windows, and plastic phenotypes that are able to track this variation will be favored – so maybe cumulative temp will be most favored, but photoperiod will be disfavored.

- Simulated variation – we should be able to create histories that independently manipulate year-to-year variation independently of day-to-day variation.

3 Underlying model

Organisms collect fitness for a number of days (default 10) starting *lag* days after they decide to emerge (default is one day). In the simulation model, a set population size for the next generation is populated using a multinomial distribution with probabilities based on the relative fitnesses of each parent. Fitness is a function of the climate, with the default function being the product of two gaussians – one based on rainfall, the other on temperature, centered on biologically realistic "optimal" values.

Organisms decide to emerge based on a linear combination of their individual traits and environmental cues. Each individual has a set of traits, corresponding to cues, which can be any real number. Every day the organism calculates their 'emergence metric' E by the linear combination of cues and traits:

$$E = (day) \ b.day \ + \ (daily \ temp) \ b.temp + \ldots$$

where the traits like *b.day* can differ between individuals, and the cues like *day* are based on that day's cues. The first day that the organism experiences an E value above 100, the organism decides to emerge. It then emerges *lag* days later (default is one), and begins collecting fitness

Table 1: Table title goes here

	Low day-to-day variation	High day-to-day variation
High year-to-year variation	- Instantaneous measures favored - Cumulative measures should be okay - Photoperiod disfavored due to high year-to-year variation	- Extreme case is totally random and should be neutral - Cumulative measures favored - Photoperiod disfavored due to high year-to-year variation - Instantaneous traits disfavored due to high day-to-day variation.
Low year-to-year variation	In the extreme case, all cues should be informative (neutral case)	- Photoperiod should be favored - Cumulative measures should be okay - Instantaneous traits disfavored due to high day-to-day variation

The table caption goes here

for *duration* days (default is 10). Organisms sum up the fitness across all days emerged to get their total fitness.

For simplicity of other operations (mainly unbiased mutation rates), we actually used a slightly different formulation of the traits, such that our linear combination was

$$E = (day)\frac{100}{b.day} + (daily\ temp)\frac{100}{b.temp} + \ldots$$

In this interpretation, a trait value represents the value of a cue necessary for the organism to emerge if it only used that cue. For example, an organism with a *b.day* trait value of 65 and all other traits set to zero would emerge on the 65th day.

Cues (which have corresponding traits) include *day* (which is a proxy for photoperiod), *temperature*, *precipitation*, *cumulative temperature*, *cumulative precipitation*, and the square of each of those (we have focused on the non-squared traits).

4   Climate

We took two approaches to choosing the climates our organisms experience.

4.1   Natural approach

In the 'natural' approach, we took climate data from weather stations that provided many years of daily data. Mostly we have focused on 100 years of daily weather data from Davis. As some years have missing measurements on individual days, we used a system of imputations to fill in missing data points in a reasonable way. With these complete years of data, we generated simulated climates by resampling the entire set of years, or some subset of years with properties of interest (variable or consistent years, for example).

4.2 Constructed approach

For situations where we wanted more control, we created a baseline by taking the the daily mean temperature and precipitation from the 100 years of davis data. We then smoothed this, ensuring that the smoothing "wrapped around" such that the last day of the year connected smoothly to the first day. From this smooth baseline year, we generated climatic regimes with more or less day-to-day variance (by adding to each day a random value of temperature and/or precipitation drawn from a normal distribution with mean 0 and variance of our choice) and year-to-year variance (by offsetting the phenology by a number drawn from a normal distribution with mean 0 and variance of our choice, subsequently rounded).

It occurs to me we could consider also giving each year a baseline temperature offset - one year might have all temperatures increased by 5 degrees, the next year decreased by 3, or what have you. This could be interesting, but seems like its effect would be largely dependent on our choice of optimal temperature/precipitation for the fitness function, where as the other sources of variance should be fairly robust

5 Analysis

We have taken two approaches to using our model to answer questions.

5.1 Simulation

We generate a climate by the methods described in 4.1 or 4.2. We generate a starting population of size N, assigning random trait values within the given range, for whichever traits the organisms are allowed to use. We then simulate the population through time, determining offspring of each generation based on their parent's fitness (assuming clonal reproduction) (parents are selected using a multinomial distribution to keep the population size at N), and add mutations randomly (occurrence of mutation is random and based on given mutation probabilities, size of mutation is random and based on given mutation variances). Because of the inherent stochasticity of this approach, we carry out multiple runs with the same parameter values to look at the general patterns across many runs.

5.2 Fitness optimization

When we're interested in what the optimal trait combination is for a given climate, we can solve this explicitly by optimizing the allowed traits for the maximum the geometric fitness. This is dramatically faster and does a better job of finding the optimum, but sacrifices the realism of mechanistically simulating biological processes. Currently our optimization is implemented by testing the geometric fitness in a large number of points in traitspace, and using the best 10-20 as starting points for runs of R's optim() function with fairly loose criterion. The final results of the best 5-10 of those optim() runs are used as the starting points for more stringent optim() runs, and we take the resulting locations in traitspace to be near to the actual optimal point.

6 Validation

We tested our simulation on some key baseline climates:

- Completely random fitness: traits appear to drift

- Completely static years: organisms rapidly begin to emerge at the optimal time of year, relying on any trait that can provide that information.

I believe we have also compared simulation results with fitness optimization results, and found that they matched. But that may have just been fitness optimization results with heatmaps of fitness through traitspace.

7 Examples of citation using bibtex

In