

# Modeling masting

From individual tree behavior to population-level synchrony

## 1 The motivations behind the project

### 💡 The ecological motivation

Understanding the reproductive behavior that arises at the population level requires to study individual trees' responses. We believe that integrating reproductive biology (i.e. biological **constraints**) and environmental **cues** at the *individual* level can help us investigating the drivers of the variability and synchronicity of reproduction at the *population* scale. Our final goal is to understand to what extent climate change could really disrupt forest regeneration—and provide better forecasts!

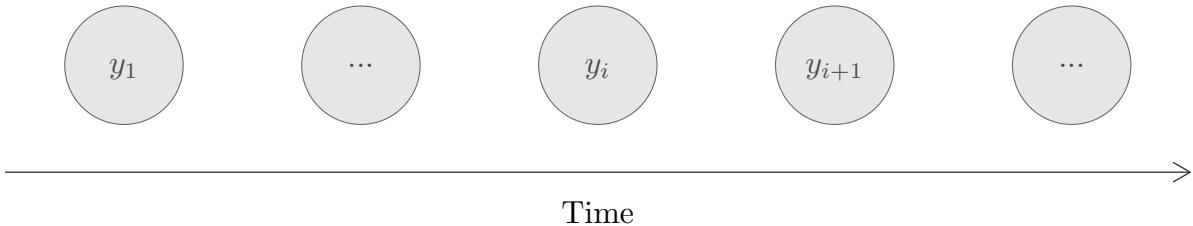
### 💡 The technical motivation

Bayesian magic! Our approach models the (latent) reproductive states of each *individual* tree, and the consequent amount of seed production. The states explicitly encode the **constraints** that shape tree reproduction – in particular, the fact that most trees need at least two years between flower bud differentiation and fruit maturation. We do not standardize away the complexities of individual-level reproduction (e.g. with normalized stand-level indices between 0 and 1...). On the contrary, modeling tree-level reproduction allows us to obtain *population* estimates – and thus direct inferences on the population-level variability and synchronicity! And the cherry on top: we incorporate climatic **cues** at different key moments of the reproductive cycle.

## 2 The model

### 2.1 Previous models and limitations

The observations we have—whether for an individual tree or a seed trap—are a time series of yearly seed counts  $y_i$ .



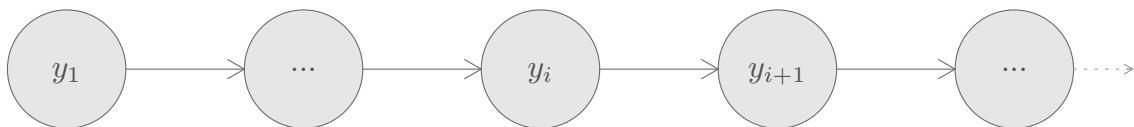
The following quote illustrates the current approach of statistical model of masting:

“We fitted a zero-inflated, negative binomial mixed model to the annual number of initiated seeds in each tree, with fixed factors that included summer temperatures in 1 and 2 years before seedfall, [...] and seed production in the previous year to account for possible resource depletion. [...] We included [...] a first-order temporal autocorrelation structure.”

If we ignore the zero-inflated part, we could write this model as something like:

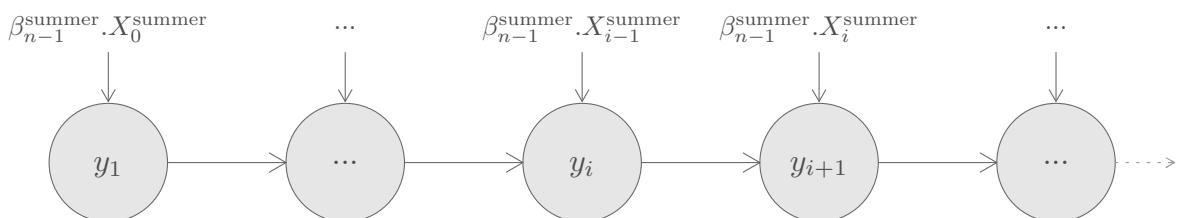
$$\begin{aligned} y_i &\sim \text{NegBin}(\mu_i, \phi) \\ \mu_i &= \alpha + \beta_{n-1}^{\text{summer}} \cdot X_{i-1}^{\text{summer}} + \beta_{n-2}^{\text{summer}} \cdot X_{i-2}^{\text{summer}} + \beta_{n-1}^{\text{seed}} \cdot y_{i-1} + \rho \cdot \epsilon_{i-1} + \epsilon_i \\ \epsilon_i &\sim \mathcal{N}(0, 1) \end{aligned}$$

The idea is to introduce some dependency between the seed counts:



The inclusion of both a lagged response of  $\beta_{n-1}^{\text{seed}} \cdot y_{i-1}$  and a first-order autocorrelation structure on the residuals  $\rho \cdot \epsilon_{i-1}$  seems... quite tricky.

But more importantly, the effect of summer temperature does not directly depends on the previous state. In other words, whatever the value of  $\beta_{n-1}^{\text{seed}} \cdot y_{i-1}$ , a tree that experiences a warm summers would be predicted to have an increasing seed production—regardless of the previous reproductive state.

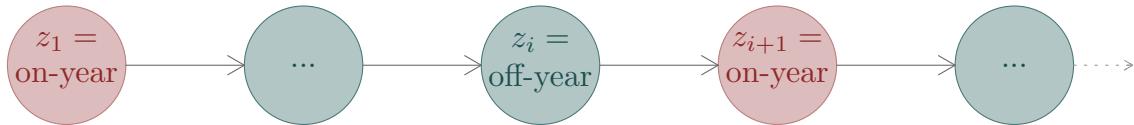


This issue arises because we do not explicitly simulate the reproductive state of an individual tree. The model does not allow to differentiate effects of climate during a non-masting year or a masting year. Moreover, without modeling individual-level states, it is not possible to infer reproductive behavior at the population scale—including the degree of potential synchrony.

## 2.2 A new approach grounded in biology

Researchers have worked on the reproductive biology of trees for decades—and in particular fruit trees. In parallel, people working on masting have defined characteristics of masting and developed clear hypotheses about the evolutionary mechanisms that could cause masting. There is thus a great opportunity to develop more generative models, that explicitly incorporate biological knowledge of flower and fruit development in order to better understand how masting works.

To understand the reproductive behavior that arises at the population level, we need to model individual trees' reproduction. At the tree level, floral buds are initiated the year before flowering—in the same time as fruits of the current year start developing. During a large crop year (an “on”-year), the presence of many fruits depress flower initiation because of hormonal inhibition and resource trade-off. Thus, the next year will likely be an “off”-year. This behavior is called **alternate bearing**. It is not always a 2-year cycle, as an “on”-year can be followed by several “off”-years. This clearly defines two states for a tree.



Each state  $z_{i+1}$  is dependent **only on the preceding state  $z_i$**  (which is called the Markovian assumption). These states are latent, because we observe only the seed counts.

With some parameters  $\theta$ , the probability of observing some sequences of the two first states is:

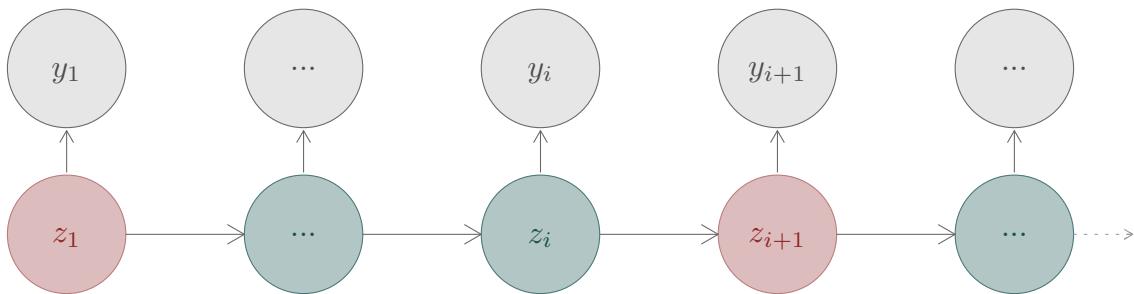
$$p(z_1, z_2 | \theta) = p(z_1) \times p(z_2 | z_1, \theta)$$

And then, adding the next state:

$$p(z_1, z_2, z_3 | \theta) = p(z_1) \times p(z_2 | z_1, \theta) \times p(z_3 | z_2, \theta)$$

And so on ...

Each observed seed count at each step arises from the state of the tree.



This gives us this likelihood for the two first seed counts:

$$p(y_1, y_2, z_1, z_2 \mid \theta) = p(z_1) p(y_1 \mid z_1, \theta) p(z_2 \mid z_1, \theta) p(y_2 \mid z_2, \theta)$$

And for all observations until  $i + 1$ :

$$p(y_1, \dots, y_{i+1}, z_1, \dots, z_{i+1} \mid \theta) = p(z_1) p(y_1 \mid z_1, \theta) \prod_{k=2}^{i+1} p(z_k \mid z_{k-1}, \theta) p(y_k \mid z_k, \theta).$$

So, what parameters do we have to estimate?

### 2.2.0.1 Transition between states

Let's look at the part which concerns the latent states.

First, we need to infer the initial state probability  $p(z_1)$ , i.e. the probability that the first year is an **on-year** or an **off-year**. Because we have only two states,  $p(z_1 = \text{on}) = 1 - p(z_1 = \text{off})$ —i.e. we have only one parameter to estimate.

We also have to infer the probabilities of transition  $p(z_k \mid z_{k-1}, \theta)$ . With two states, this corresponds to this transition matrix:

$$M = \begin{bmatrix} \tau_{\text{on} \rightarrow \text{on}} & \tau_{\text{on} \rightarrow \text{off}} \\ \tau_{\text{off} \rightarrow \text{on}} & \tau_{\text{off} \rightarrow \text{off}} \end{bmatrix}$$

The probability  $p(z_2)$  is then obtain with the product:

$$\begin{bmatrix} p(z_2 = \text{on}) & p(z_2 = \text{off}) \end{bmatrix} = \begin{bmatrix} p(z_1 = \text{on}) & p(z_1 = \text{off}) \end{bmatrix} \cdot M$$

Again, the transition matrix corresponds to only two parameters, for example:

$$M = \begin{bmatrix} \tau_{\text{on} \rightarrow \text{on}} & 1 - \tau_{\text{on} \rightarrow \text{on}} \\ \tau_{\text{off} \rightarrow \text{on}} & 1 - \tau_{\text{off} \rightarrow \text{on}} \end{bmatrix}$$

In summary, we have 3 parameters to infer: the initial probability of masting  $\tau_0 = p(z_1 = \text{on})$ , the probability of transitioning from an **off-year** to an **on-year**  $\tau_{\text{off} \rightarrow \text{on}}$ , and the probability of staying into an **on-year** state  $\tau_{\text{on} \rightarrow \text{on}}$ .

### 2.2.0.2 Seed production parameters

Now, what about the parameters of the observational model that links the latent state to the observed seed count?