HMM model on abundance classes

In practice, information about standing flora is available in terms of abundance classes, not counts. It is also more realistic to work with abundance classes of seeds in the bank. We therefore extended the first HMM model based on counts to a model based on abundance classes. We defined $c_y$ and $c_x$, the abundance classes of seeds and of standing plants observed, respectively. Since no information is available on the distribution of the counts of seeds/emerged plants within a given class, we decided to attach a uniform distribution to each class on the interval defining the class. For the maximal abundance class that is only lower bounded by 101, we attached a geometric distribution of parameter $p = 1/101$ to this class. In our HMM model, the abundance class of mature plants, $c_x^{t+1}$, observed between $t$ and $t+1$, only depends on the abundance class of seeds, $c_y^t$, at the end of the previous year and on $a'$, the agricultural practices applied between $t$ and $t+1$. The abundance class of seeds, $c_y^{t+1}$, depends on $c_y^t$, the abundance class of seeds, $c_x^{t+1}$, the observed abundance class of plants between $t$ and $t+1$, and $a'$, the practices applied between $t$ and $t+1$ (Fig. 1 of the main paper).

Hence, the probability that a weed species reaches an abundance class $c_x^{t+1}$ at $t+1$ where $c_y^t$ is its abundance class at $t$, and $a'$ is the set of agricultural practices applied between $t$ and $t+1$, is:

$$P_{a'}(c_x^{t+1} \mid c_y^t) = \sum_{x^{t+1} \in I_{c_x^{t+1}}} \sum_{y' \in I_{c_y^t}} P(y' \mid c_y^t) P_{a'}(x^{t+1} \mid y')$$  \hspace{1cm} (eq. A2)

where $I_{c_x^{t+1}}$ and $I_{c_y^t}$ represent the intervals of values in abundance classes $c_x^{t+1}$ and $c_y^t$, respectively.
The probability that a weed species population reaches the seed abundance class \( c_{y^{t+1}} \) in the seed bank at \( t+1 \) given its abundance of seeds \( c_{y'} \) at \( t \) in the seed bank and its abundance of standing plants \( c_{x^{t+1}} \) at \( t+1 \) where \( a' \) is the set of agricultural practices applied between \( t \) and \( t+1 \), is:

\[
P_{\alpha'}(c_{y^{t+1}} | c_{x^{t+1}}, c_{y'}) = \sum_{y^{t+1} \in I_{c_{y'}}} \sum_{y' \in I_{c_{y'}}} \sum_{x^{t+1} \in I_{c_{x'}}} \sum_{m \in I_{c_{m}}} P_{\alpha'}(y_{m}^{t+1} | y', x^{t+1}) P(y'_{m} | c_{y'}) \ldots P(x^{t+1} | c_{x'}) P_{\alpha'}(Y_{p}^{t+1} = y^{t+1} - y' + x^{t+1} + y_{m}^{t+1} | x^{t+1})
\]

(eq. A3)

Transition probabilities were estimated by simulation. In order to compute \( P_{\alpha'}(c_{x^{t+1}} | c_{y'}) \) for each of the six classes \( c_{y'} \), \( K \) random realisations of \( Y' \) were generated, following the distribution of \( c_{y'} \). For each random realisation of \( Y' \), random realisations of \( X^{t+1} \) were generated, following the germination rate (\( \sigma \)) of \( a' \), and attached to the corresponding classes. The same method was used to compute \( P_{\alpha'}(c_{y^{t+1}} | c_{x^{t+1}}, c_{y'}) \). The accuracy of this computation can be controlled by the parameter \( K \). For small values of \( K \), computation is faster but less accurate than with a large value of \( K \).

In the abundance class model defined above (eq. A2 and A3), we assumed that abundance classes are defined as intervals of counts, ignoring possible errors in counting. However, this assumption is unrealistic, especially when counts are close to abundance class boundaries. In order to take this potential counting error into account, an error rate (\( \varepsilon \)) was included in the computation of transition probabilities. It represents the probability of an error of abundance class estimation of +/- one class (for non-extreme abundance classes). This error rate (\( \varepsilon \)) was fixed to 0.2.