



HAL
open science

Detecting seed bank influence on plant metapopulation dynamics

Apolline Louvet, Nathalie Machon, Jean-baptiste Mihoub, Alexandre Robert

► **To cite this version:**

Apolline Louvet, Nathalie Machon, Jean-baptiste Mihoub, Alexandre Robert. Detecting seed bank influence on plant metapopulation dynamics. *Methods in Ecology and Evolution*, 2021, 12 (4), pp.655-664. 10.1111/2041-210x.13547 . hal-03202810

HAL Id: hal-03202810

<https://hal.science/hal-03202810>

Submitted on 20 Apr 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

1 **This is the pre-peer reviewed version of the following article:** Louvet, A.,
2 Machon, N., Mihoub, J. B., Robert, A. (2021). Detecting seed bank influence on plant
3 metapopulation dynamics. *Methods in Ecology and Evolution*, 12(4), 655-664., **which**
4 **has been published in final form at <https://doi.org/10.1111/2041-210X.13547>.**
5 **This article may be used for non-commercial purposes in accordance with**
6 **Wiley Terms and Conditions for Use of Self-Archived Versions.**

7 Detecting seed bank influence on plant
8 metapopulation dynamics

9 Apolline Louvet¹, Nathalie Machon², Jean-Baptiste Mihoub², Alexandre Robert²

10 ¹ Centre de Mathématiques Appliquées (CMAP), Ecole Polytechnique, Centre Na-
11 tional de la Recherche Scientifique, Institut Polytechnique de Paris, Route de Saclay,
12 91128 Palaiseau Cedex, France

13 ² Centre d'Ecologie et des Sciences de la Conservation (CESCO), Muséum National
14 d'Histoire Naturelle, Centre National de la Recherche Scientifique, Sorbonne Université,
15 Paris, France

16
17 Corresponding author : Apolline Louvet (apolline.louvet@polytechnique.edu)

18
19 Running headline : Seed bank detection in plant metapopulations

Abstract

1. Seed banks are known to play a key role in plant metapopulations. However, detecting seed banks remains challenging and requires intense monitoring efforts. Assessing the genuine effect of seed banks on plant metapopulation dynamics may offer a much easier while still biologically relevant way to overcome this issue.

2. In this study, we developed a new metric : the Seed Bank Characteristic Event (SBCE) probability. Instead of detecting seed bank directly, the SBCE probability measures seed bank contribution to the observed metapopulation dynamics. Exploring seed bank parameters (colonization, germination and seed bank death probabilities, initial proportion of patches containing a seed bank), a wide range of monitoring durations (from 3 to 10 years) and number of patches in the metapopulation (from 10 to 1000 patches), we examined the conditions under which the SBCE probability is correctly estimated. To test the robustness of our approach, we further introduced false negatives, false positives or parameter heterogeneity between patches. Finally, we applied the SBCE probability method to the monitoring of tree bases plant species in Paris, France, to assess the applicability of the method to real-world datasets. We studied the influence of species traits and environmental characteristics, in order to increase understanding of plant metapopulation dynamics within an urban environment.

3. Our results indicate that the SBCE probability is well estimated when enough monitoring years or number of patches are considered, and for rates of false negatives or false positives of up to 0.1. However, the SBCE probability estimation is not robust to colonization probability heterogeneity between patches. When we applied the SBCE probability method to the real monitoring dataset, we found a weak influence of the flowering months of the species and of the green space the closest to the metapopulation on SBCE probability estimates.

4. The study suggests that the measurement of seed bank contribution is less data-demanding than assessment of seed bank presence. Applying the es-

51 timation method to the monitoring of tree bases plant species highlights a
52 significant contribution of the seed bank to plant metapopulation dynamics
53 in an urban environment, and illustrates how the method can be applied on
54 real-world datasets.

55 **Keywords :** Hidden Markov Model, metapopulation, plants, Propagule Rain model,
56 seed bank, urban biodiversity, tree bases

57 Introduction

58 An important issue in ecology and conservation biology is determining the mechanisms
59 underlying persistence of plant or animal populations in fragmented landscapes (Fahrig,
60 2003). In plants, the seed bank, i.e the spontaneous storage of seeds within the soil, plays
61 a critical role in metapopulation and community dynamics (Fenner, 2017). However,
62 assessing directly the presence of a seed bank, for instance by putting soil samples in
63 germination chambers, and measuring its associated parameters is challenging. There-
64 fore conceptual approaches and statistical tools allowing one to estimate these quantities
65 using widespread data such as presence/absence data can prove very useful. One suitable
66 conceptual framework for studying patchy environments is the metapopulation theory,
67 first introduced in (Levins, 1969). A metapopulation is defined as a population living
68 in a set of patches that can be colonized or go extinct, the regional persistence of the
69 species resulting from a balance between local colonizations and extinctions (MacArthur
70 & Wilson, 1967). Statistical tools have been developed to allow parameter inference for a
71 broad range of metapopulation models (see e.g (Moilanen, 1999, 2004)), and were fruit-
72 fully used in studies on insects (Hanski, 2011; Moilanen, Smith, & Hanski, 1998) or small
73 mammals (Ozgul, Armitage, Blumstein, & Oli, 2006).

74 As plants form populations with a strong spatial structure and can only move from one
75 patch to another as propagules, metapopulation models appear at first as particularly
76 suited to their study (Husband & Barrett, 1996). Yet classical metapopulation models do
77 not account for seed banks, which are common in seed plants (Baskin & Baskin, 2014),
78 potentially leading to erroneous estimates of extinction and colonization rates (Fréville,
79 Choquet, Pradel, & Cheptou, 2013) and making these models generally irrelevant for
80 studying plant metapopulation dynamics (Freckleton & Watkinson, 2002). New models
81 taking into account the influence of a seed bank were developed recently (Fréville et al.,
82 2013; Borgy, Reboud, Peyrard, Sabbadin, & Gaba, 2015). These models consider the seed
83 bank state as an *hidden state*, which is not visible but which influences patch occupancy,
84 and which can be estimated from patch occupancy data. In this article, we elaborated a
85 new method to characterize seed bank contribution to metapopulation dynamics based

86 on the model introduced in (Pluntz et al., 2018). This model allows parameter infer-
87 ence on a variant with a seed bank of a classical model, the *Propagule Rain Model* (or
88 PRM) (Gotelli, 1991), in which patches are colonized or go extinct independently from
89 each other, with the same colonization or extinction probability. However, a limit of
90 the model proposed by (Pluntz et al., 2018) is that it would not be applicable in many
91 real-world situations. It indeed requires either a long monitoring duration or several
92 thousand patches to be monitored in order to accurately estimate all parameters (that
93 is, germination, colonization and seed bank death probabilities). To overcome this prob-
94 lem, we introduced in the present study a new metric more accurately estimated than
95 the metapopulation parameters for identical monitoring duration and number of patches,
96 providing information about the influence of the seed bank on plant population dynamics
97 in real populations. Then, we used this metric with both theoretical and real metapopu-
98 lation data. We called this metric the *seed bank characteristic event* probability (SBCE
99 probability). As it only measures the contribution of the seed bank to the observed stand-
100 ing vegetation dynamics, it is less informative than knowing all seed bank parameters,
101 but the goal of this study was to show that it is accurate in more real-life situations.
102 We performed analyses based on simulated presence/absence time series data, and on
103 time series data in which we introduced some flaws commonly found in real datasets, in
104 order to give bounds on the number of patches or of years of observation needed to fulfill
105 different accuracy requirements. Our goal was to provide guidelines on how to design the
106 data collection step.

107 As a case study, we used the estimation method on annual floristic inventories of nat-
108 ural and spontaneous flora carried out from 2009 to 2018 on 1324 tree bases located in
109 Paris, France (the *Paris 12* dataset). Indeed, the population of plants in urban tree
110 bases is located inside an inhospitable matrix and has an high turnover, which makes
111 metapopulation models particularly suited (Dornier, Pons, & Cheptou, 2011). Studies
112 using presence/absence data for various species present in urban tree bases considered
113 as metapopulations were already carried out, but to our knowledge none accounted for
114 seed bank potential presence in tree bases (Omar et al., 2019; Dornier et al., 2011). We

115 also attempted to relate the contribution of the seed bank to the observed population
116 dynamics to species traits and environmental characteristics. The results are intended to
117 increase the understanding of population dynamics inside an urban environment.

118 Overall, our study (i) gives insights on the importance of seed bank contribution to plant
119 metapopulation dynamics within an urban environment, and (ii) provides a compre-
120 hensive framework to detect the effects of seed banks in plant metapopulations, which can
121 be applied on a wide range of ecological systems, including but not restricted to urban
122 environments.

123 **Material and methods**

124 **Model used**

125 The model from (Pluntz et al., 2018) we used in this study is a variant of the Propagule
126 Rain Model (PRM) (Gotelli, 1991). In the PRM, colonization and extinction probabilities
127 do not depend on the current state of the metapopulation, and are constant over patches
128 and time. A seed bank can be introduced using Hidden Markov Model (HMM) techniques
129 (Cappé, Moulines, & Rydén, 2005; Rabiner, 1989). The seed bank contains the seeds that
130 were just produced by standing vegetation of the focal patch or came from colonization
131 events by the propagule rain, along with seeds produced by previous generations that
132 did not germinate yet and are still alive. Since all plants originate from the seed bank
133 of the patch they are in, the presence of plants at one time step means that the seed
134 bank contained seeds just before germination could occur. Therefore metapopulation
135 parameters can be estimated along with the presence/absence of seeds in the seed bank
136 at each time-step.

137 The model is characterized by these three parameters :

- 138 • the joint probability of seed germination and of survival of seedlings until adulthood.
139 This parameter will be called *germination probability*, and denoted g , following the
140 existing literature.

- 141 • the *colonization probability* c of the patch by external seeds entering the seed bank.
- 142 • the *seed bank extinction probability conditional on the seed bank not having germinated* d . This parameter is the probability that the seed bank will not survive until
143 the next generation, assuming it has not germinated yet.
144

145 The initial proportion p_0 of patches containing a seed bank can be considered as an extra
146 parameter of the model.

147 The model evolves as follows : for each patch, if the seed bank is not empty, seeds can
148 germinate with probability g . If they do, the plants will grow and produce seeds which
149 will refill the seed bank. Otherwise, the seed bank can survive until the next generation
150 with probability $1-d$. New seeds can enter the seed bank during a colonization event with
151 probability c independently of the presence of standing vegetation or seeds in any given
152 patch (see figure 1). The main difference with the PRM model is that if the seed bank
153 death probability d is strictly less than 1, then germination can be delayed. Hereafter,
154 models for which $d < 1$ will be denoted as SB+ (*Seed-Bank Plus*) and those for which
155 $d = 1$, corresponding to the classical PRM, will be denoted SB- (*Seed-Bank Minus*).

156 The procedure for estimating parameters uses the Expectation-Maximization (EM) al-
157 gorithm (Dempster, Laird, & Rubin, 1977) in order to find the best parameter fit (see
158 details in (Pluntz et al., 2018)). We set the number of iterations in the algorithm to
159 200 in order to ensure convergence of parameters. As the EM algorithm can converge
160 to a local but non-global maximum, the choice of the initial conditions can affect the
161 value returned by the algorithm. We preliminary checked with simulated datasets that
162 convergence of the EM algorithm to a non-global maximum was rather unlikely (results
163 not shown).

164 The algorithm was first used to get theoretical results on the performances of the SBCE
165 estimation method. We simulated datasets for different parameter values for g , c , d and
166 p_0 and applied the SBCE estimation method on these simulated datasets in order to iden-
167 tify conditions under which the SBCE probability offers reliable estimates using a set of
168 performance criteria. Then, we used the estimation method on a real monitoring dataset,
169 the Paris 12 dataset (see below). Due to the structure of this dataset, we implemented a

170 variant of the estimation method supporting one year of missing data.

171 **Criterion for seed bank identification : SBCE probability**

172 Seed Bank Characteristic Event (SBCE) probability, denoted \mathbb{P}_{SBCE} , is defined as the
173 probability for standing flora to produce seeds that (1) will stay in the seed bank during
174 more than one year (without germinating nor dying), and (2) will germinate before new
175 seeds come from an external source. In other words, \mathbb{P}_{SBCE} is the probability that
176 standing flora in a given patch at a given time-step contributes with delay to population
177 dynamics, and that this contribution is not hidden by colonization events happening in
178 between. This probability can be computed knowing g , c and d . Denoting SB_t the event
179 *during t years, the seed bank does not germinate, nor receive new seeds from external*
180 *source, nor dies*, then :

$$\begin{aligned}\mathbb{P}_{SBCE} &= \sum_{t \geq 1} \mathbb{P}(SB_t) \times g \\ &= g \times \sum_{t \geq 1} \mathbb{P}(SB_1)^t \\ &= g \times \sum_{t \geq 1} [(1-g)(1-c)(1-d)]^t \\ &= g \times \frac{(1-g)(1-c)(1-d)}{1 - (1-g)(1-c)(1-d)}\end{aligned}$$

181 SBCE probability can be interpreted as a measure of seed bank contribution to metapop-
182 ulation dynamics : it takes into account that some parameter values, e.g high colonization
183 or seed bank extinction probabilities, make it hard for delayed germination events to be
184 observed. As a result it cannot be used per se to conclude to seed bank presence or
185 absence, but is a proxy of seed bank contribution to the metapopulation dynamics. A
186 consequence is that low SBCE probability does not mean that there is no seed bank,
187 since a high colonization probability yields comparable effects.

188 Hereafter, we will consider the effect of the seed bank to be :

- 189 • *weak* if SBCE probability is lower than 0.05

- 190 • *medium* if SBCE probability is in the interval $[0.05, 0.10]$
- 191 • *strong* if SBCE probability is higher than 0.10.

192 These thresholds were defined arbitrarily using values taken by the SBCE probability in
 193 situations where the observed effects of the seed bank can be considered weak, strong or
 194 intermediate. Continuous values taken by the SBCE probability for parameter values in
 195 table 1 are presented in Supplementary Information (A.1). An estimator of SBCE proba-
 196 bility must satisfy two requirements so that it can accurately show seed bank contribution
 197 to metapopulation dynamics. First, the estimated SBCE probability must be accurately
 198 estimated for datasets generated with a SB+ model. We considered the estimation was
 199 correct if the root-mean-square error (RMSE) on SBCE probability estimation was lower
 200 than the threshold of 0.14 used in (Pluntz et al., 2018). Then, it must avoid, as much as
 201 possible, identification of medium or strong seed bank effects in datasets generated with
 202 a SB- model, for which $\mathbb{P}_{SBCE} = 0$. In this article, the rate of false identification of a
 203 medium to strong seed bank effect in a model without seed bank (SB-), denoted $p_{falseSB}$
 204 thereafter, is deemed satisfying if it is below 0.05.

205 In order to assess the performance of the criterium, we investigated the sensibility of
 206 \mathbb{P}_{SBCE} estimation to c , g , d and p_0 by building 24 distinct parameter sets combining a
 207 broad range of values for these parameters (16 corresponding to SB+ models, and 8 to
 208 SB- models, see table 1). Each parameter set was used to generate 30 datasets for each
 209 of the 4 monitoring durations and 8 numbers of patches listed in table 1. We chose to
 210 consider durations of at most 20 years and at most 1000 patches.

211 We computed the estimated SBCE pobability by performing parameter fits of c , g , d
 212 and p_0 on the simulated datasets. The accuracy of the SBCE probability estimation was
 213 tested separately for each SB+ parameter set, time duration and number of patches by
 214 computing the RMSE. We determined the minimal number of patches needed to sat-
 215 isfy a RMSE threshold of 0.14 (as in (Pluntz et al., 2018)). For SB- parameter sets, as
 216 $\mathbb{P}_{SBCE} = 0$, we instead computed $p_{falseSB}$ for each combination of monitoring duration
 217 and number of patches.

218 The performance of the criterium were compared to the one of a criterium comparing

219 the AIC of parameter fits with or without a constraint on d (i.e, comparing the AIC
220 of parameter fits on a SB+ and a SB- model). The AIC-based criterium was deemed
221 satisfying if it was able to identify the correct model at least 25 times out of 30 (i.e, 83%
222 of correct assignments).

223 **Testing SBCE estimation robustness**

224 Before applying the parameter estimation method to real data, we tested the algorithm
225 robustness to several flaws commonly found in real datasets : false negatives, false posi-
226 tives and parameter heterogeneity.

227 Parameter heterogeneity was only studied for colonization probability by assuming that
228 a proportion of patches has a colonization probability equal to 0. Details of the protocols
229 are provided in Supporting Information (A.2).

230 **Applying SBCE probability to real-world monitoring data**

231 The real dataset used in this study to apply the SBCE probability, Paris 12, consists of
232 floristic inventories of 1324 tree bases located in Paris 12th administrative district, car-
233 ried out annually between 2009 and 2018. Natural and spontaneous flora was inventoried
234 exhaustively over the entire period, except in 2013, when a limited number of species
235 were tracked. For those species that were not monitored in 2013, we used the missing
236 data variant of the model for the year 2013. The taxonomic reference is the French Flora
237 Reference TAXREF v8.0 (Gargominy et al., 2014). See Supporting Information (B.1) or
238 (Omar, Al Sayed, Barré, Halwani, & Machon, 2018) for information about nomenclature
239 and the species and streets monitored.

240 For almost every species, a high proportion of patches were never occupied from 2009
241 to 2018. We interpreted this as being due to colonization heterogeneity, and considered
242 that patches which were never colonized had a colonization probability equal to 0. Since
243 the germination and seed bank death probabilities cannot be estimated for a group of
244 patches that were never occupied, we removed these patches from the analysis.

245 We considered that the tree bases present in different streets represented distinct metapop-

246 ulations. Moreover, we considered that the metapopulation parameters could be different
247 from one species to another, and for a given species, from one street (i.e one metapop-
248 ulation) to another. Each pair of species and street was analyzed separately, and was
249 retained only if the species was observed at least once in more than 20 distinct tree bases
250 of the street (the size of a street ranged from 31 to 186 tree bases). For each pair, seed
251 bank contribution was assessed under the hypothesis that the species' patch occupancy
252 dynamics followed the Propagule Rain Model. Parameter estimation for SB+ model was
253 carried out using the missing data variant when the species was not inventoried in 2013.
254 We then tested whether the estimated SBCE probability was affected by the nature of the
255 closest green space (see (Omar et al., 2019)), the seed dispersal mechanism, the flowering
256 months (extracted from the database of the collaborative network of French botanists
257 "Tela botanica" (<http://www.tela-botanica.org>), the releasing height of the seeds and the
258 seed weight (mean value obtained from the LEDA database (Kleyer et al., 2008)). Be-
259 cause almost all species flower in summer months (June, July and August) and due to
260 high correlation in flowering for consecutive months, we limited the flowering period data
261 to two binary variables : early flowering (months of March, April or May) and late flow-
262 ering (September or October). We used mixed-effect linear regression model with SBCE
263 probability as the response variable, species and street identity as random factors, the
264 greenspace the closest to the street among 4 (one park, one footpath, railways and the
265 Seine river) and the dispersal mechanism (anemochorous, autochorous, barochorous or
266 epizoochorous) as fixed qualitative variables, the releasing height of the seeds and the seed
267 weight as fixed quantitative variables and the early and late flowering as binary variables,
268 assuming a Gaussian distribution. We checked whether the requirements of independence
269 of residuals, normality of residuals and homogeneity of variances were met. The model
270 was implemented with the lmer function of the lme4 R package (Bates, Mächler, Bolker,
271 & Walker, 2014). Besides the above regression model, we also performed repeatability
272 analyses to provide an overview of the variation of the germination probability and SBCE
273 probability among streets and among species. The repeatability analysis was based on
274 1000 parametric bootstraps as implemented in the rptR package of R (Stoffel, Nakagawa,

275 & Schielzeth, 2017). The statistical significance of the repeatability of each metric was
276 tested by a likelihood ratio test comparing the model fit of a model including a grouping
277 factor (here, the species or the street) and one excluding it.

278 **Results**

279 **Criterion for seed bank identification**

280 For each monitoring duration, we could find a number of patches ensuring the cri-
281 terium fulfillment (RMSE < 0.14 on SBCE probability estimation for SB+ models, and
282 $p_{falseSB} < 0.05$ for SB- models). No such number of patches could be found for the
283 criterium comparing AIC of parameter fits on a model with or without seed bank, which
284 is not based on SBCE probability (see Supporting Information (A.3)).

285 Overall, the minimal number of patches required decreased from 500 for a monitoring
286 lasting 3 years to 30 for a monitoring lasting 20 years, and was higher for SB+ models
287 than for SB- models. Complete results can be found in Supporting Information (A.1).

288 **Testing SBCE estimation robustness**

289 Overall, the introduction of false negatives, false positives or heterogeneous colonization
290 increased RMSE on SBCE probability estimation for SB+ models, and $p_{falseSB}$ for SB-
291 models.

292 For SB+ models, for all the false positive or false negative rates considered, the RMSE
293 almost always stayed below the 0.14 threshold. The introduction of heterogeneous col-
294 onization by making some patches not colonizable led to the threshold being quickly
295 exceeded, except for 500 patches and 3 years of monitoring.

296 For SB- models, when we introduced false negatives or false positives, increasing the
297 monitoring duration or the number of patches led to a decrease of $p_{falseSB}$ (albeit less
298 marked for a false negative rate of 0.2), and the 0.05 threshold could be met for a suffi-
299 cient monitoring duration or number of patches, except for a false negative rate of 0.2 (see
300 table 3). Conversely, when we introduced heterogeneous colonization (i.e, when we set

301 the colonization probability of some patches to 0), increasing the number of patches had
302 no marked effect, and increasing the monitoring duration actually made $p_{falseSB}$ increase,
303 reaching values of up to 0.5. This effect was particularly marked when considering only
304 the false identification rate of a strong (instead of a medium to strong) seed bank effect.
305 Consequently, no combination of number of years of observation and number of patches
306 monitored fulfilled the accuracy requirements, even for a proportion of non-colonizable
307 patches of 0.05.

308 Complete results can be found in Supporting Information (A.4).

309 **Applying SBCE probability to real-world monitoring data**

310 The analysis highlighted a high variability of SBCE probabilities, both between species
311 and within species. A medium to strong seed bank effect was detected in at least one
312 street for most species. We further investigated the reason for this variation looking
313 at estimates of the germination probability g . Germination probabilities were gener-
314 ally fairly low, most of the time below 0.5, no matter the species or the street. The
315 repeatability analysis indicated that the germination and SBCE probabilities were very
316 consistent between streets for a given species (repeatability $R = 0.595 \pm 0.07$ for ger-
317 mination and $R = 0.38 \pm 0.08$ for SBCE, p-value $< 10^{-4}$ for both repeatabilities). In
318 contrast, repeatabilities of these probabilities between species for a given street were low
319 ($R = 0.058 \pm 0.03$, p-value = 7×10^{-4} for germination and $R = 0.058 \pm 0.04$, p-value = 0.003
320 for SBCE). Overall, these results suggest that both germination and SBCE probabilities
321 primarily depend on species rather than location. Results are summarized in figures 2
322 and 3. Estimations of SBCE and germination probabilities for each pair of species and
323 street (i.e, each metapopulation) analysed can be found in Supporting Information (B.2).
324 According to the regression model, none of our explaining variables was significantly
325 correlated with the SBCE probability (see detailed regression results in Supporting In-
326 formation (B.2)). However, the type of the nearest green space and the flowering period
327 exhibited marginally significant effects : late flowering and being closest to the park were
328 associated with higher SBCE probability (figure 3).

329 Discussion

330 In this paper, we propose a new metric, the seed bank characteristic event probability
331 (SBCE) providing information on the contribution of the seed bank to the observed
332 standing vegetation dynamics in plant metapopulations. Our results indicated that the
333 SBCE performs well in a wide range of situations and provided evidence of a significant
334 contribution of the seed bank to plant metapopulation dynamics in an urban environment.
335 In biology, as in other disciplines, it is sometimes more practical and straightforward
336 to make inferences about a process within a system by observing the effects of that
337 process on the system rather than the process itself. This approach is central to the
338 study of metapopulations, where a process such as dispersal is often studied indirectly by
339 examining its consequences in terms of genetic structuring or recolonization dynamics.
340 This idea is also at the core of the analytical framework recently developed ([Fréville et](#)
341 [al., 2013](#); [Borgy et al., 2015](#)) to study seed banks, which are very difficult to detect on a
342 large scale, but whose consequences in terms of plant metapopulation dynamics can be
343 crucial ([Fenner, 2017](#)).

344 Theoretical analysis

345 Our analysis indicates that the SBCE approach is relevant to evaluate the contribution
346 of a potential seed bank to the dynamics of a metapopulation in which patches are in-
347 dependent, i.e., in which the colonization or extinction of a patch does not depend on
348 the presence of the species in the other patches, even when the number of monitored
349 patches and monitoring duration are limited. In particular, we showed that for limited
350 monitoring durations and number of patches, the SBCE probability provides more precise
351 predictions than model identification methods using AIC and based on estimates of the
352 germination probability g , the colonization probability c and the seed bank death prob-
353 ability d . This could be expected, since even though the SBCE probability is computed
354 using g , c and d , it depends mostly on g , which according to ([Pluntz et al., 2018](#)) is well
355 estimated. Conversely, model estimation methods depend more on estimates of d , which

356 is more difficult to estimate. For a sufficient but still achievable number of patches or of
357 years of observations, the SBCE probability was well estimated when seed bank existed
358 and correctly reflected the absence of a seed bank otherwise (see Supporting Informa-
359 tion (A.1)). As this metric measures the contribution of the seed bank to the global
360 metapopulation dynamics, it gives information on whether the seed bank is essential to
361 the observed dynamics. However, it is not suited for getting estimates of the seed bank
362 parameters, namely probabilities of germination and survival of seeds (but see (Pluntz et
363 al., 2018)).

364 Our work highlights that false negatives or false positives in standing vegetation detection
365 and heterogeneous colonization have contrasted impacts on the estimation of the SBCE
366 probability. On the one hand, for models without seed bank, increasing false positive rate
367 or false negative rate lead to an increase of the rate of false identification of a medium to
368 strong seed bank effect $p_{falseSB}$. The presence of false positives can be mitigated by mon-
369 itoring more patches over a longer duration, while the presence of false negatives can be
370 mitigated only when the false negative rate is not too high. This implies that monitoring
371 methods reducing the false negative rate may increase the accuracy of the SBCE prob-
372 ability estimate, even if they are traded-off by an increase of the rate of false positives.
373 On the other hand, with strongly heterogeneous colonization (e.g., in the special case in
374 which a fraction of patches can never be colonized), the SBCE probability cannot be well
375 estimated, no matter the monitoring duration or the number of patches. Therefore it
376 is crucial to identify heterogeneous colonization situations, whose impact on estimation
377 accuracy cannot be mitigated by improving the monitoring effort. Other statistical meth-
378 ods can be used in order to identify heterogeneous colonization cases, allowing to treat
379 the case where colonization probabilities are heterogeneous but potentially all non-zero,
380 for instance using mixture models (Robin, 2018). However these methods are far more
381 computationally intensive than the simple one we used. The consistency of estimates of
382 the probability of germination g between streets for most species in the *Paris 12* dataset
383 suggests that the method we introduced is still efficient.

384 The fact that increasing the number of patches can mitigate the impact of false negatives

385 or false positives suggests that the method may be used on data coming from citizen
386 science programs, for which the number of patches monitored can be very large. Recent
387 empirical results on citizen science programs showed that, with a standardized proto-
388 col and good training methods, error rates in datasets can reach values lower than 0.05
389 (Fuccillo, Crimmins, de Rivera, & Elder, 2015; Ratnieks et al., 2016), which makes these
390 programs compatible with our SBCE approach.

391 **The Paris 12 dataset**

392 The analysis performed on the *Paris 12* dataset of plants present in tree bases in Paris
393 showed that 55 % of the pairs of species and streets analysed exhibited a medium to strong
394 seed bank effect, suggesting that seed banks have a key influence on plant metapopulation
395 dynamics in this type of urban environment. Moreover, estimates of both germination
396 probability g and SBCE probability are consistent between streets for a given species,
397 but they are not consistent from one species to another for a given street. Besides, the
398 spatial variation of SBCE estimates among streets is weakly related to which green space
399 is the closest to the street. Explaining this pattern is difficult, as it may result from the
400 green space characteristics, from the design of the tree bases or the tree-base management
401 performed by the green space services of the city of Paris. Most estimates of g are low,
402 below 0.5 or even 0.3, which means that either germination per se is low, either survival of
403 seedlings until adulthood is low. Potential explanations are a strong competition between
404 species, or a removal of plants before seed production by gardeners.

405 Our analysis does not uncover any influence of the dispersal mechanism on the SBCE
406 probability. This result is in line with the findings of (Omar et al., 2018), who showed
407 that the distribution of species in Paris tree bases was not correlated to the dispersal
408 mechanism nor to the weight of the seeds, and hypothesized that this was partly due to
409 human activity spreading all seeds no matter the weight or the dispersal device (Sukopp,
410 2004; Von der Lippe & Kowarik, 2007).

411 Our study also highlights a weak influence of the flowering period on the SBCE proba-
412 bility. Species whose flowering period starts early (before May) exhibited a higher mean

413 estimated SBCE probability than those whose flowering period starts later. One possi-
414 ble interpretation is that plant species that flower early are more detectable than those
415 flowering late, so they are more likely to be removed by gardeners, which affects the
416 probability of survival of seedlings until adulthood and the colonization probability.

417 The applicability of the SBCE probability on the *Paris 12* dataset has some limitations.
418 First, according to our theoretical analysis, estimations of the SBCE probability for a
419 monitoring duration of 10 years is expected to be accurate if at least 100 tree base sare
420 considered, and more in case of false negatives or false positives. As streets in our dataset
421 have between 30 and 150 tree bases and as some tree bases were removed from analysis
422 to avoid issues related to heterogeneous colonization, small errors on SBCE probability
423 estimates have to be expected, and particularly a slight overestimation of SBCE proba-
424 bilities when no seed bank is present (i.e when SBCE probability is equal to 0).

425 Second, we assumed that tree bases' dynamics are independent of each other (i.e., no
426 colonization from one patch to another). As a result, we did not consider the alternative
427 hypothesis of Levins metapopulation model ([Levins, 1969](#)) with a seed bank. In this
428 model, colonization events do not bring seeds from a propagule rain, but from neigh-
429 bouring patches. Therefore, colonization probability is not the same for all patches,
430 and depends on the state of the metapopulation. An estimation method for this model
431 exists ([Le Coz, Cheptou, & Peyrard, 2019](#)), but uses abundance data instead of pres-
432 ence/absence data, and would need to be adapted to handle missing data. Moreover, the
433 fact that we recovered results previously documented with a different approach and the
434 consistency of g estimates for a given species between streets suggests that the assump-
435 tion of independent patches was correct.

436 Our results show that measuring seed bank contribution to plant metapopulation dynam-
437 ics is less data-demanding than assessing seed bank presence, while being robust to the
438 presence of false negatives or false positives. Our method can be applied to a wide range
439 of urban and non-urban metapopulations, and can be used on datasets collected using cit-
440 izen science in order to increase substantially the understanding of plant metapopulation
441 dynamics.

442 **Acknowledgments**

443 We wish to thank Mona Omar, Marion Noualhaguet, Marion Dubois, Ambre Zéléla
444 Bouvard, Paul Haenel, Noëlie Maurel, Florence Devers, Hélène Beaugeard, Sébastien
445 Julliard, Laure Schneider-Maunoury and Gwendoline Chastel who carried out the floristic
446 survey from 2009 to 2018. This work was supported by ANR (Agence Nationale de la
447 Recherche) Écoville [ANR 14 CE22-0021] and was done during the internship of AL at
448 the CESCO laboratory as a student of ENS Paris.

449 **Authors' contributions**

450 All authors conceived the ideas and AL designed methodology. AL analyzed the data
451 and led the writing of the manuscript. All authors contributed critically to the drafts
452 and gave final approval for publication.

453 **Data availability**

454 The *Paris 12* dataset is available on Zenodo (<http://doi.org/10.5281/zenodo.3770339>).

References

- 455
- 456 Baskin, C., & Baskin, J. (2014). *Seeds : Ecology, biogeography, and evolution of dormancy*
457 *and germination*. Academic Press, San Diego.
- 458 Bates, D., Mächler, M., Bolker, B., & Walker, S. (2014). Fitting linear mixed-effects
459 models using lme4. *arXiv preprint arXiv:1406.5823*.
- 460 Borgy, B., Reboud, X., Peyrard, N., Sabbadin, R., & Gaba, S. (2015). Dynamics of
461 weeds in the soil seed bank: a hidden markov model to estimate life history traits
462 from standing plant time series. *PloS one*, *10*(10), e0139278. doi: [10.1371/jour-](https://doi.org/10.1371/journal.pone.0139278)
463 [nal.pone.0139278](https://doi.org/10.1371/journal.pone.0139278)
- 464 Cappé, O., Moulines, E., & Rydén, T. (2005). *Inference in hidden markov models*.
465 Springer-Verlag, New York.
- 466 Dempster, A. P., Laird, N. M., & Rubin, D. B. (1977). Maximum likelihood from
467 incomplete data via the em algorithm. *Journal of the royal statistical society. Series*
468 *B (methodological)*, 1–38. doi: [10.1111/j.2517-6161.1977.tb01600.x](https://doi.org/10.1111/j.2517-6161.1977.tb01600.x)
- 469 Dornier, A., Pons, V., & Cheptou, P.-O. (2011). Colonization and extinction dynamics
470 of an annual plant metapopulation in an urban environment. *Oikos*, *120*(8), 1240–
471 1246. doi: [10.1111/j.1600-0706.2010.18959.x](https://doi.org/10.1111/j.1600-0706.2010.18959.x)
- 472 Fahrig, L. (2003). Effects of habitat fragmentation on biodiversity. *Annual re-*
473 *view of ecology, evolution, and systematics*, *34*(1), 487–515. doi: [10.1146/an-](https://doi.org/10.1146/annurev.ecolsys.34.011802.132419)
474 [nurev.ecolsys.34.011802.132419](https://doi.org/10.1146/annurev.ecolsys.34.011802.132419)
- 475 Fenner, M. (2017). Ecology of seed banks. In *Seed development and germination* (pp.
476 507–528). Routledge.
- 477 Freckleton, R., & Watkinson, A. (2002). Large-scale spatial dynamics of plants: metapop-
478 ulations, regional ensembles and patchy populations. *Journal of Ecology*, *90*(3),
479 419–434. doi: [10.1046/j.1365-2745.2002.00692.x](https://doi.org/10.1046/j.1365-2745.2002.00692.x)
- 480 Fréville, H., Choquet, R., Pradel, R., & Cheptou, P.-O. (2013). Inferring seed bank
481 from hidden markov models: new insights into metapopulation dynamics in plants.
482 *Journal of Ecology*, *101*(6), 1572–1580. doi: [10.1111/1365-2745.12141](https://doi.org/10.1111/1365-2745.12141)
- 483 Fuccillo, K. K., Crimmins, T. M., de Rivera, C. E., & Elder, T. S. (2015). Assessing

484 accuracy in citizen science-based plant phenology monitoring. *International journal*
485 *of biometeorology*, 59(7), 917–926. doi: [10.1007/s00484-014-0892-7](https://doi.org/10.1007/s00484-014-0892-7)

486 Gargominy, O., Terceirie, S., Régnier, C., Ramage, T., Schoelinck, C., Dupont, P., ...
487 Poncet, L. (2014). Taxref v8. 0, référentiel taxonomique pour la france: Méthodolo-
488 gie, mise en oeuvre et diffusion. *Rapport SPN*, 42, 2014.

489 Gotelli, N. J. (1991). Metapopulation models: the rescue effect, the propagule rain,
490 and the core-satellite hypothesis. *The American Naturalist*, 138(3), 768–776. doi:
491 [10.1086/285249](https://doi.org/10.1086/285249)

492 Hanski, I. A. (2011). Eco-evolutionary spatial dynamics in the glanville fritillary butterfly.
493 *Proceedings of the National Academy of Sciences*, 108(35), 14397–14404. doi:
494 <https://doi.org/10.1073/pnas.1110020108>

495 Husband, B. C., & Barrett, S. C. (1996). A metapopulation perspective in plant popu-
496 lation biology. *Journal of Ecology*, 461–469. doi: [10.2307/2261207](https://doi.org/10.2307/2261207)

497 Kleyer, M., Bekker, R., Knevel, I., Bakker, J., Thompson, K., Sonnenschein, M., ...
498 others (2008). The leda traitbase: a database of life-history traits of the north-
499 west european flora. *Journal of Ecology*, 96(6), 1266–1274. doi: [10.1111/j.1365-](https://doi.org/10.1111/j.1365-2745.2008.01430.x)
500 [2745.2008.01430.x](https://doi.org/10.1111/j.1365-2745.2008.01430.x)

501 Le Coz, S., Cheptou, P.-O., & Peyrard, N. (2019). A spatial markovian framework for
502 estimating regional and local dynamics of annual plants with dormancy. *Theoretical*
503 *population biology*, 127, 120–132. doi: [10.1016/j.tpb.2019.03.002](https://doi.org/10.1016/j.tpb.2019.03.002)

504 Levins, R. (1969). Some demographic and genetic consequences of environmental het-
505 erogeneity for biological control. *American Entomologist*, 15(3), 237–240. doi:
506 [10.1093/besa/15.3.237](https://doi.org/10.1093/besa/15.3.237)

507 MacArthur, R. H., & Wilson, E. O. (1967). *The theory of island biogeography*. Princeton
508 university press, New Jersey.

509 Moilanen, A. (1999). Patch occupancy models of metapopulation dynamics: efficient
510 parameter estimation using implicit statistical inference. *Ecology*, 80(3), 1031–
511 1043. doi: [10.1890/0012-9658\(1999\)080\[1031:POMOMD\]2.0.CO;2](https://doi.org/10.1890/0012-9658(1999)080[1031:POMOMD]2.0.CO;2)

512 Moilanen, A. (2004). Spomsim: software for stochastic patch occupancy mod-

513 els of metapopulation dynamics. *Ecological modelling*, 179(4), 533–550. doi:
514 [10.1016/j.ecolmodel.2004.04.019](https://doi.org/10.1016/j.ecolmodel.2004.04.019)

515 Moilanen, A., Smith, A. T., & Hanski, I. (1998). Long-term dynamics in a metapop-
516 ulation of the american pika. *The American Naturalist*, 152(4), 530–542. doi:
517 [10.1086/286188](https://doi.org/10.1086/286188)

518 Omar, M., Al Sayed, N., Barré, K., Halwani, J., & Machon, N. (2018). Drivers of the
519 distribution of spontaneous plant communities and species within urban tree bases.
520 *Urban Forestry & Urban Greening*, 35, 174–191. doi: [10.1016/j.ufug.2018.08.018](https://doi.org/10.1016/j.ufug.2018.08.018)

521 Omar, M., Schneider-Maunoury, L., Barré, K., Al Sayed, N., Halwani, J., & Machon, N.
522 (2019). Colonization and extinction dynamics among the plant species at tree bases
523 in paris (france). *Ecology and Evolution*, 9(15), 8414–8428. doi: [10.1002/ece3.4954](https://doi.org/10.1002/ece3.4954)

524 Ozgul, A., Armitage, K. B., Blumstein, D. T., & Oli, M. K. (2006). Spa-
525 tiotemporal variation in survival rates: implications for population dynamics
526 of yellow-bellied marmots. *Ecology*, 87(4), 1027–1037. doi: [10.1890/0012-9658\(2006\)87\[1027:SVISRI\]2.0.CO;2](https://doi.org/10.1890/0012-9658(2006)87[1027:SVISRI]2.0.CO;2)

528 Pluntz, M., Coz, S. L., Peyrard, N., Pradel, R., Choquet, R., & Cheptou, P.-O. (2018).
529 A general method for estimating seed dormancy and colonisation in annual plants
530 from the observation of existing flora. *Ecology letters*. doi: [10.1111/ele.13097](https://doi.org/10.1111/ele.13097)

531 Rabiner, L. R. (1989). A tutorial on hidden markov models and selected applications in
532 speech recognition. *Proceedings of the IEEE*, 77(2), 257–286. doi: [10.1109/5.18626](https://doi.org/10.1109/5.18626)

533 Ratnieks, F. L., Schrell, F., Sheppard, R. C., Brown, E., Bristow, O. E., & Garbuzov, M.
534 (2016). Data reliability in citizen science: learning curve and the effects of training
535 method, volunteer background and experience on identification accuracy of insects
536 visiting ivy flowers. *Methods in Ecology and Evolution*, 7(10), 1226–1235. doi:
537 [10.1111/2041-210X.12581](https://doi.org/10.1111/2041-210X.12581)

538 Robin, S. (2018, January). *Models with hidden structure with applications in biology*
539 *and genomics*. Master 2 Mathématiques pour les Sciences du Vivant, Université
540 Paris-Sud.

541 Stoffel, M. A., Nakagawa, S., & Schielzeth, H. (2017). rptR: Repeatability estimation

- 542 and variance decomposition by generalized linear mixed-effects models. *Methods in*
543 *Ecology and Evolution*, 8(11), 1639–1644.
- 544 Sukopp, H. (2004). Human-caused impact on preserved vegetation. *Landscape and urban*
545 *planning*, 68(4), 347–355. doi: [10.1016/S0169-2046\(03\)00152-X](https://doi.org/10.1016/S0169-2046(03)00152-X)
- 546 Von der Lippe, M., & Kowarik, I. (2007). Long-distance dispersal of plants by vehi-
547 cles as a driver of plant invasions. *Conservation Biology*, 21(4), 986–996. doi:
548 [10.1111/j.1523-1739.2007.00722.x](https://doi.org/10.1111/j.1523-1739.2007.00722.x)

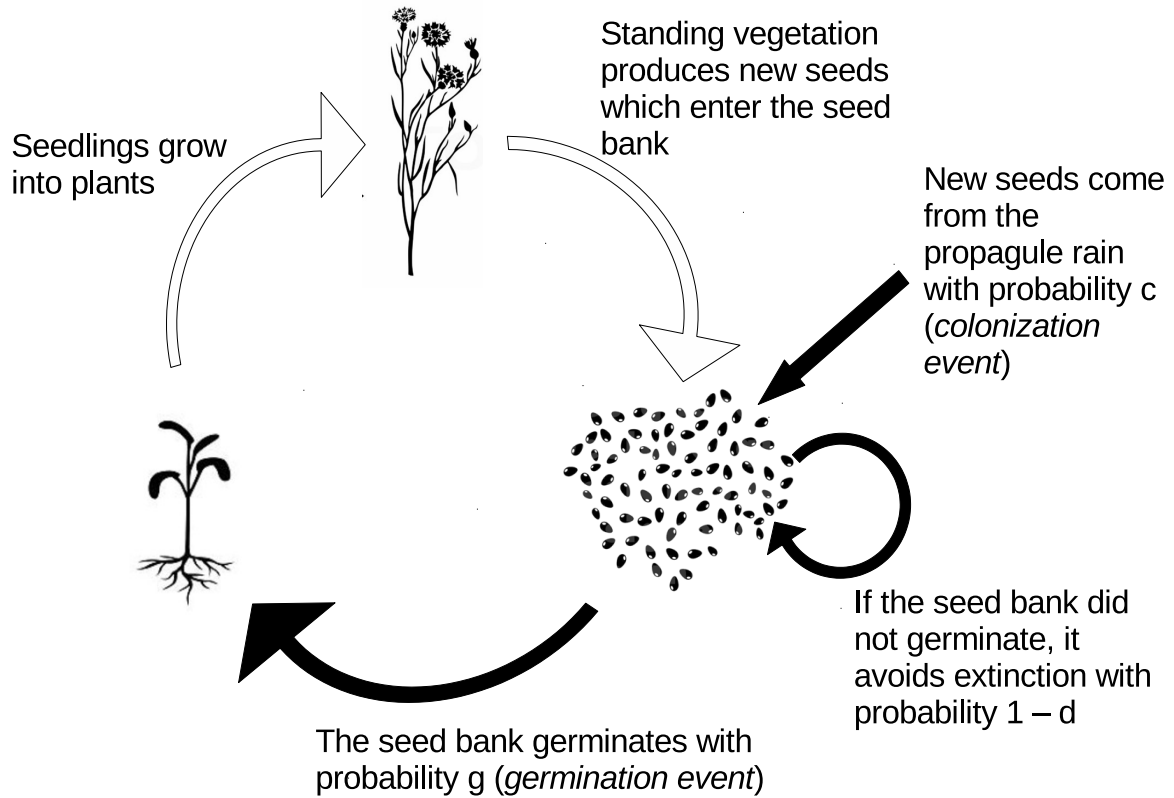
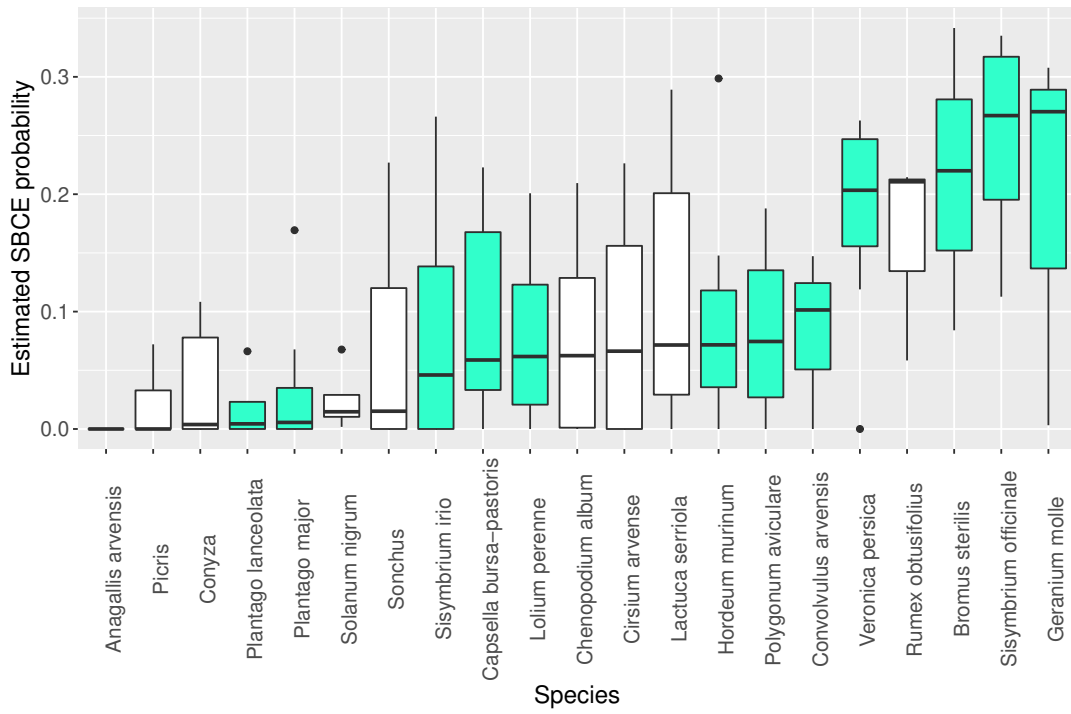


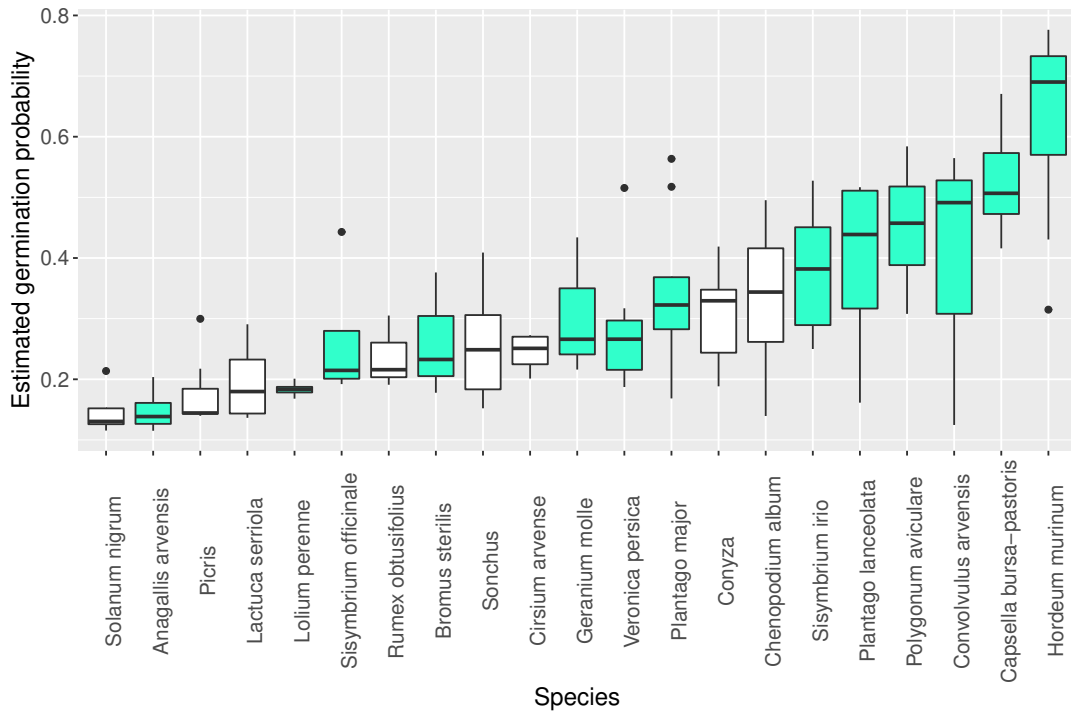
Figure 1: Graphical representation of the variant with seed bank of the Propagule Rain Model. White arrows indicate transitions that always occur, while black arrows indicate transitions that only occur with a fixed probability.

Parameter and notation	SB+ model	SB- model
Colonization probability (c)	0.3, 0.7	0.3, 0.7
Germination probability (g)	0.3, 0.7	0.3, 0.7
Seed bank death probability (d)	0.2, 0.6	1
Initial proportion of seeds (p_0)	0.3, 0.7	0.3, 0.7
Years of observation	3, 5, 10, 20	3, 5, 10, 20
Number of patches	10, 30, 50, 100, 200, 500, 800, 1000	10, 30, 50, 100, 200, 500, 800, 1000

Table 1: Parameter sets and test conditions used for testing the seed bank identification criterium.



(a) SBCE probability estimates



(b) Germination probability estimates

Figure 2: (a) Distribution of SBCE probability estimates for species present in at least 20 tree bases of at least 3 different streets. (b) Distribution of germination probability estimates for species present in at least 20 tree bases of at least 3 different streets. Species flowering early are indicated in turquoise, and species flowering late in white. Species whose flowering period could not be found on Tela Botanica are not indicated.

Monitoring duration	3 years		5 years		10 years		20 years	
Model	SB+	SB-	SB+	SB-	SB+	SB-	SB+	SB-
Minimal number of patches	500	200	200	100	100	100	30	30

Table 2: Minimal number of patches needed to achieve $RMSE < 0.14$ on SBCE probability estimation for SB+ models, and $p_{falseSB} < 0.05$ for SB- models, depending on monitoring duration.

Monitoring duration	3 years	5 years	10 years	20 years
False negative rate	Number of patches			
0	500	200	100	30
0.05	1000	500	500	200
0.1	X	1000	500	500
0.2	X	X	X	X
False positive rate	Number of patches			
0	500	200	100	30
0.05	X	1000	500	100
0.1	X	X	500	500
0.2	X	X	X	800

Table 3: Minimal number of patches needed to satisfy both $RMSE < 0.14$ on SBCE probability estimation for SB+ models and $p_{falseSB} < 0.05$, after having introduced false negatives or false positives in the dataset, for different monitoring durations and rates of false negatives or false positives. The symbol X means that criterium satisfaction could not be met.

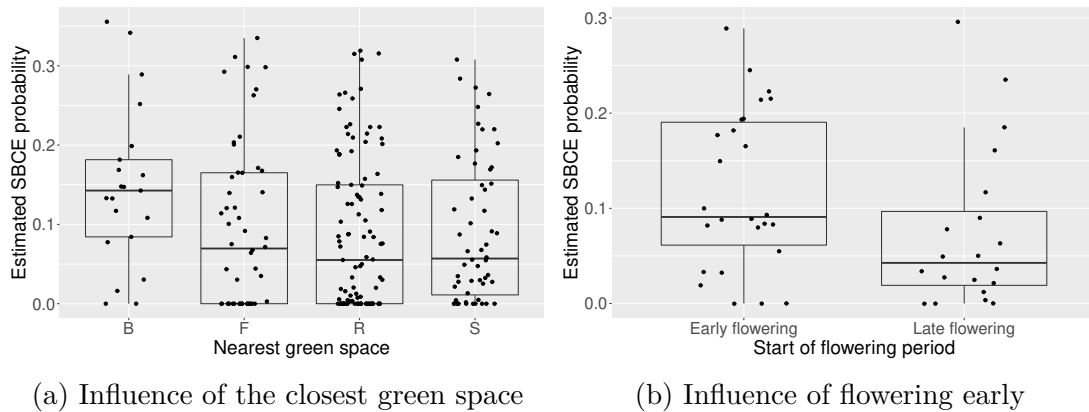


Figure 3: (a) Distribution of SBCE probability estimates in streets classified by the nature of the closest green space (see Supplementary information (B.1) for green spaces names). (b) Distribution of SBCE probability estimates for species with early flowering period (March, April or May), compared to those with late flowering period.