HAL
archives-ouvertes.fr

Inference for Partially Observed Multitype Branching Processes and Ecological Applications

Catherine Laredo, Olivier David, Aurélie Garnier

▶ **To cite this version:**

Catherine Laredo, Olivier David, Aurélie Garnier. Inference for Partially Observed Multitype Branching Processes and Ecological Applications. 31 pages, 1 figure. 2008. <hal-00363957>

HAL Id: hal-00363957
https://hal.archives-ouvertes.fr/hal-00363957

Submitted on 24 Feb 2009

INFERENCE FOR PARTIALLY OBSERVED MULTITYPE BRANCHING PROCESSES AND ECOLOGICAL APPLICATIONS

By Catherine Larédo * , Olivier David ,
and Aurélie Garnier'

INRA, Jouy-en-Josas and PMA Universités Paris 6 & 7

Multitype branching processes with immigration in one type are used to model the dynamics of stage-structured plant populations. Parametric inference is first carried out when count data of all types are observed. Statistical identifiability is proved together with derivation of consistent and asymptotically Gaussian estimators for all the parameters ruling the population dynamics model. However, for many ecological data, some stages (i.e. types) cannot be observed in practice. We study which mechanisms can still be estimated given the model and the data available in this context. Parametric inference is investigated in the case of Poisson distributions. We prove that identifiability holds for only a subset of the parameter set depending on the number of generations observed, together with consistent and asymptotic properties of estimators. Finally, simulations are performed to study the behaviour of the estimators when the model is no longer Poisson. Quite good results are obtained for a large class of models with distributions having mean and variance within the same order of magnitude, leading to some stability results with respect to the Poisson assumption.

1. Introduction. Understanding population dynamics requires models that admit the complexity of natural populations and the data ecologists can get from them. Thus analyzing ecological data raises questions ranging from modeling purposes to statistical inference. Among various methods, Leslie matrices or demographic matrix models are widely used for studying the dynamics of age or stage-structured populations (e.g. Caswell, 2001). These models are deterministic with noise added to introduce some variability. In many cases however and especially for small populations, the

---

demographic stochasticity has to be taken into account; these models are too simple (Melbourne and Hastings, 2008) and can no longer be used, even adding stochasticity into the dynamics with random effects and covariates (Royle, 2004; Barry et al., 2003) or Bayesian approaches (Raftery et al., 1995; Gross et al., 2002; Clark and Bjornstad, 2004). For these reasons, we use here stochastic models to study small populations dynamics.

The starting point of this work is a three-year field survey of feral populations (i.e. populations escaped from crops) of an annual crop species (oilseed rape) that was carrried out in the center of France (Selommes, Loir-et-Cher; Garnier, 2006). Unlike cultivated oilseed rape, very few facts are known about the dynamics of feral oilseed rape populations. In this study, the dynamics is modeled by a multitype branching process (five types including vegetative and reproductive plant stages along with seeds in the soil seed-bank) with immigration in one type (seeds). Data consisted in populations counts in each type, except the seeds that could not be observed. Three main difficulties occur when studying this demographic dataset. (1) A large number of populations ($K = 300$) have been observed over a short period of time ($n = 2, 3$); (2) only count data have been collected; (3) some types could not be observed by ecologists (here seeds). These characteristics are clearly not specific to this survey and are frequently met in data coming from Population Genetics and Ecology (see e.g. de Valpine, 2004 and the references therein). These data could be studied as longitudinal data, but for concerns about the dynamics, better insights can be obtained by means of mechanistic models describing it.

Branching processes have largely been studied (see Athreya and Ney, 1972 for a general presentation; Mode, 1971 for Multitype Branching Processes and Haccou et al., 2005; Kimmel and Axelrod, 2002; Mode and Sleeman, 2000 for applications in biology). Statistical inference has also been largely investigated (Hall and Heyde, 1980; Guttorp, 1991 for general branching processes; Wei and Winnicki, 1990; Winnicki, 1991 for branching processes with immigration; Bhat and Adke, 1981; Maaouia and Touati, 2005; Gonzalez et al., 2008 for multitype branching processes). However, the precise multitype branching process with immigration used here is a combination of the previous ones and moreover statistical inference for multitype branching processes is usually based on the following different observations (Maaouia and Touati, 2005; Gonzalez et al., 2008): the number of descendants of type j coming from all the type i individuals. We just observed the successive counts of "individuals" of each type. This is more realistic assumption since this situation frequently occurs with datasets from field studies, and inference is studied here in this framework.

We are interested in the estimation of the parameters involved in the population dynamics from the incomplete observations of count data collected simultaneously in several populations. This is an "Incomplete Data model", or "State Space model" (as defined for instance in (Cappé et al., 2005). It is also an inverse problem and a central theme in Ecology arising from its study is that parameters might not be identifiable knowing only the population dynamics (Wood, 1997). In practice, the inference based on such data is performed using various E.M. algorithms eventually coupled with Monte Carlo methods (Dempster et al., 1977; Kuhn and Lavielle, 2004; McLachlan and Krishnan, 2007; Sung and Geyer, 2007; Olsson et al., 2008), and Bayesian or Hierarchical Bayesian methods (Clark and Bjornstad, 2004; Buckland et al., 2004; Thomas et al., 2005). All these methods circumvent but cannot address the identifiability problem. However, identifiability is a prerequisite of statistical inference, and understanding the dynamics mechanisms strongly relies on how parameters are linked in the identifiability question. We propose here an integrated framework in order to analyze as accurately as possible the whole data set of the field survey. Introduction of covariates and a priori knowledge, errors coming from non exhaustive population samplings within some populations, use of various algorithms rely on this work and are studied in two companion papers (David et al., 2008; Garnier et al., 2008).

The paper is organized as follows. Section 2 contains the description of the population dynamics and preliminary results (Proposition 2.1). The statistical inference for complete observations is studied in Section 3. We first prove identifiability for all the parameters and derive consistent and asymptotically Gaussian estimators (Proposition 3.1). Since seeds are not observed in practice, the problem of unobserved types is addressed in Section 4. This is a non linear non Gaussian state space model. The associated three-dimensional stochastic process is no longer Markovian. The model with Poisson distributions provides a useful example with explicit computations. We obtain a closed form of the dependence of present observations on the whole past for the three-dimensional process (Theorem 4.1). A question concerns the statistical model identifiability : it is studied according to the number of observed generations. We characterize the parameter subset where identifiability holds (Theorem 4.2) and study the parametric inference (Proposition 4.2). Section 5 presents simulation results to study how the estimation performs with respect to deviations to the Poisson model. Detailed proofs are given in the Appendix.

2. Model and preliminary results.

2.1. *Dynamics of annual plants.* We consider annual plants with the following life cycle. Seeds are released at the end of summer; they can either enter in a seed bank if buried or germinate in autumn.The emerged rosettes vernalize during winter, then bolt in spring and finally produce mature plants that shed seeds in summer and then die. Five developmental stages are considered: rosettes before winter R, rosettes after winter (vernalized rosettes) V, mature plants F, seeds located in the soil seed bank ("old seeds") S, and seeds located on the soil surface ("new seeds") T. New seeds and old seeds are separated because they have different demographic parameters. Within each cycle, new seeds can enter these populations at the end of summer (immigration). There exist two sources of seed immigration, seeds from adjacent mature crops and seeds from spillage during seed transport (Crawley and Brown, 1995; Claessen et al., 2005a; Garnier et al., 2006; Pivard et al., 2008).
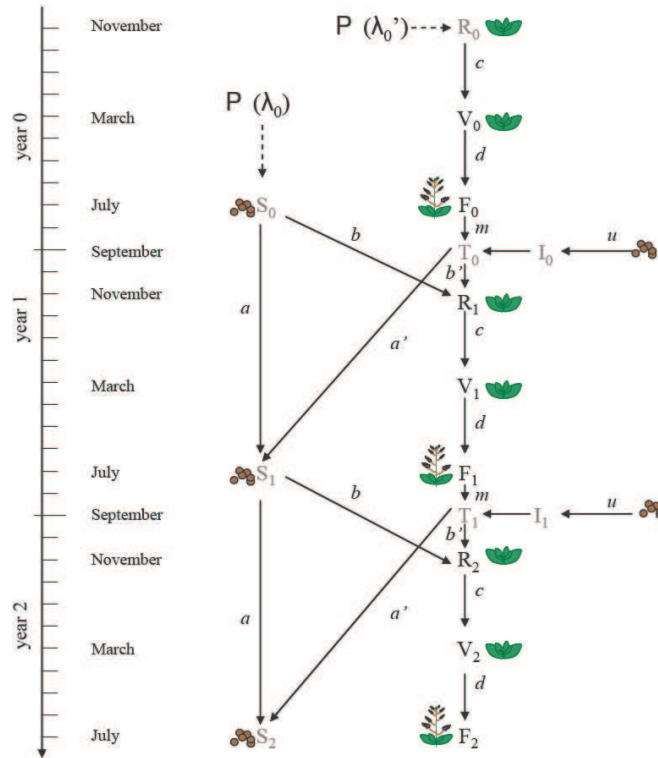

Fɪɢ 1. *Schematic Dynamics of feral oilseed rape populations.*

This model is quite general and for dynamical purposes only, it could be simplified considering just seeds and mature plants. However, our concern is different since we aim at estimating as many parameters as possible given both the model and the data available. Keeping a five-type model allows using all the data collected in the field survey and thus leads to the best description of the plant population dynamics for inference.

2.2. *Notations and assumptions.* Consider first one population. From now on, the term year corresponds to one life cycle. It starts with the birth of the new seeds and ends just before the birth of the new seeds of the next generation. All the variables are integer random variables indexed by $i \in \mathbb{N}$. For year i, denote by S_i the number of "old seeds", T_i the number of "new seeds", R_i the number of rosettes before winter, V_i the number of rosettes after winter and F_i the number of mature plants. Six parameters describe these transitions:

$$(2.1) \qquad (c, d, a, b, a', b') \in (0, 1)^6 \text{ with } 0 < a + b \leq 1, \ 0 < a' + b' \leq 1.$$

$$(2.2) \qquad P(\text{seed in } S_i \to S_{i+1}) = a \ ; \ P(\text{seed in } S_i \to R_i) = b$$

$$(2.3) \qquad P(\text{seed in } T_i \to S_{i+1}) = a' \ ; \ P(\text{seed in } T_i \to R_i) = b'$$

$$(2.4) \qquad P(\text{rosette in } R_i \to V_i) = c \ ; \ P(\text{rosette in } V_i \to F_i) = d.$$

Mature plants in F_i produce "new seeds" T'_{i+1} according to the offspring distribution $G(.)$.

A number I_i of "new seeds" immigrate into the population at the beginning of year i; it is assumed to follow the distribution $\mu(.)$. Seeds in S_i come from two sources, the "old seeds" S_{i-1} and the "new seeds" T_{i-1}. Denote by S'_i and S''_i these two quantities. Similarly, rosettes before winter R_i come from "old seeds" in S_i and "new seeds" in T_i. Denote by R'_i and R''_i the rosettes coming from S_i and T_i. These variables satisfy :

$$(2.5) \qquad S_i = S'_i + S''_i \ , \ R_i = R'_i + R''_i \ ,$$

$$(2.6) \qquad T_i = T'_i + I_i \ .$$

Note that the probabilities of dying for stages S, T, R, V are respectively $(1 - a - b)$, $(1 - a' - b')$, $(1 - c)$, $(1 - d)$ and that the probability of no offspring for a mature plant is $G(0)$.

Let us now detail the framework and assumptions used in the sequel. The field survey consisted of a large number of feral oilseed rape populations (around 300 in Garnier, 2006) observed over a short period of time ($n = 2$ or 3). These populations were isolated, so we assume here independence for these populations and the plant density was low in the surveyed populations so that density-dependence in plant survival and reproduction could be neglected (Pivard et al., 2008; Garnier, 2006). Moreover, we assume

ASSUMPTION 1. *Offspring distribution $G(.)$*
All mature plants reproduce independently according to the same offspring distribution $G(.)$ with expectation m and variance δ^2.

ASSUMPTION 2. *There is no competition in survival and germination between seeds in the seed bank or "old seeds" S_i and seeds on the soil or "new seeds" T_i.*

ASSUMPTION 3. *Immigration distribution $\mu(.)$*
(i) Immigration I_i is independent of seed bank seeds S_i, offspring seeds T_i' and of previous years.
(ii) The random variables $(I_i, i \in N)$ are independent and identically distributed according to $\mu(.)$ with expectation u and variance ρ^2.

The above assumptions are twofold: "independence" and "identically distributed". They do not have the same status : while releasing the "independence " assumption is quite difficult, the "identically distributed " assumption is done here for sake of clarity. The independence assumption in (A1) is justified because of the low plant density. There is no biological background for considering competition between the evolution of "old seeds" S_i and "new seeds" T_i' leading to (A2). In the field survey of feral oilseed rape populations, seed immigration mainly occured from spillage during seed transport and from adjacent mature crops, leading to the independence assumptions in (A3) (Pivard et al., 2008; Garnier, 2006). Adding covariates or a priori information can easily be introduced within this framework, which amounts to remove the "identically distributed" assumption. This indeed has been done for the statistical analysis on the whole data set: using that many populations had been observed, covariates, a priori knowledge, and a dependence with respect to i, k of the parameters defined in (2.2)-(2.4) have been added (Garnier et al., 2008). The framework detailed here is therefore quite general.

2.3. *Preliminary results.* From now on, time i is associated with a complete cycle of a plant. Let us still consider one population. The model is a discrete time stochastic process (X_i) with state space $\mathbb{N}^5$. Set

$$(2.7) \qquad X_i = (S_i, T_i, R_i, V_i, F_i) \text{ and } \mathcal{F}_i = \sigma(X_0, X_1, \ldots, , X_i).$$

For $x = (s, t, r, v, f)$, $x' = (s', t', r', v', f') \in \mathbb{N}^5$, denote $\pi_0(x)$ the initial distribution and $p(x, x')$ the conditional distribution of X_1 given X_0 :

$$(2.8) \qquad \pi_0(x) = \mathbb{P}(X_0 = x) \; ; \; p(x, x') = \mathbb{P}(X_1 = x'/X_0 = x).$$

For the Binomial distribution $\mathcal{B}(N; p)$, we write $\mathbb{P}(Y = k) = \mathcal{B}(N; p)(k)$; Multinomial distributions on $\mathbb{N}^l$, $\mathcal{M}(N; p_1, \ldots, p_l)$ are simplified omitting the last component, which leads for $l = 3$ and $0 < p_1 + p_2 < 1$,

$$(2.9) \quad \mathcal{M}(N; p_1, p_2)(n_1, n_2) = \mathcal{M}(N; p_1, p_2, 1 - p_1 - p_2)(n_1, n_2, N - n_1 - n_2)$$

Let $\star$ denote the convolution product of two distributions. Using these notations and (2.1), define the distributions

$$(2.10) \qquad \nu(s, t) = \mathbb{P}(S_0 = s, T_0 = t),$$

$$(2.11) \qquad p_1(s'/s, t, r) = \frac{(\mathcal{M}(s; a, b) \star \mathcal{M}(t; a', b'))(s', r)}{(\mathcal{B}(s; b) \star \mathcal{B}(t; b'))(r)},$$

$$(2.12) \qquad p_2(t'/f) = (G^{\star f} \star \mu)(t'); \; p_3(r/s, t) = (\mathcal{B}(s; b) \star \mathcal{B}(t; b'))(r),$$

$$(2.13) \qquad p_4(v/r) = \mathcal{B}(r; c)(v) , \; p_5(f/v) = \mathcal{B}(v; d)(f).$$

Using now definitions (2.7), (2.8) and notations (2.9)- (2.13), the following holds.

PROPOSITION 2.1. *Under (A1), (A2), (A3), $(X_i)_{i \geq 0}$ is a time homogeneous Markov chain on $\mathbb{N}^5$ with initial distribution $\pi_0(x)$ and transition probabilities $p(x, x')$ satisfying*

$$(2.14) \qquad \pi_0(x) = \nu(s, t) \, p_3(r/s, t) \, p_4(v/r) \, p_5(f/v)$$
$$(2.15) \qquad p(x, x') = p_1(s'/s, t, r) \, p_2(t'/f) \, p_3(r'/s', t') \, p_4(v'/r') \, p_5(f'/v').$$

The process $(X_i)_{i \geq 0}$ is also a multitype branching process with immigration in one type.

The last statement of Proposition 2.1 is immediate since, considering each stage of the plant as a type, each plant reproduces independently from the others according to the same offspring distribution with values in $\mathbb{N}^5$. However, the two types R_i and V_i have no offspring in the next generation, leading to a non positively regular multitype process as defined in Mode (1971) or Athreya and Ney (1972). The process (S_i, T_i, F_i) is positively regular process but, for the reasons stated in 2.1, we prefer keeping here the five-dimensional process (X_i).

The proof of Proposition 2.1 is given in Appendix A.1.

3. Likelihood and inference for complete observations. We first study the case when all types are observed.

3.1. *Notations and statistical framework.* We assume in the sequel that the initial distribution of (S_0, R_0), the offspring distribution $G(.)$ and the immigration $\mu(.)$ belong to the parametric families :
- distribution of (S_0, T_0) : $(\nu(\theta^1; s, t), \; \theta^1 \in \Theta^1)$;
- offspring $G(.)$: $(G(\theta^2; .) \;, \; \theta^2 \in \Theta^2)$ with mean m and variance δ^2;
- immigration $\mu(.)$: $(\mu(\theta^3; .), \; \theta^3 \in \Theta^3)$ with mean u and variance ρ^2.

Let us denote by $\theta = (c, d, a, b, a', b', \theta^1, \theta^2, \theta^3)$ (resp. θ_0) an arbitrary value (resp. the true value) of the parameter and by Θ the parameter set. We assume :

ASSUMPTION 4. Θ *compact set of* $\mathbb{R}^l$ *and* $\theta_0 \in \overset{\circ}{\Theta}$.

For the kth population, $X^k = (X_i^k, i \in \mathbb{N})$ is the Markov chain describing the population dynamics and $x_i^k = (s_i^k, t_i^k, r_i^k, v_i^k, f_i^k)$ are the observations at time i. In order to simplify the expressions for the likelihood, we consider here that S_i, T_i are observed up to time $n+1$. This has no consequence since seeds are not observed in practice. Hence we denote,

$$(3.1) \qquad\qquad X_{0:n}^k = (X_0^k, \ldots, X_n^k, S_{n+1}^k, T_{n+1}^k),$$

$$(3.2) \qquad\qquad \mathcal{O}_{0:n}^k = (x_0^k, x_1^k, \ldots, x_n^k, s_{n+1}^k, t_{n+1}^k).$$

The processes $X_{0:n}^k$ are repetitions of $X_{0:n} = (X_0, \ldots, X_n, S_{n+1}, T_{n+1})$. Joining all the populations, define :

$$(3.3) \quad X_{0:n}(K) = (X_{0:n}^1, \ldots, X_{0:n}^K), \quad \text{and} \quad \mathcal{O}_{0:n}(K) = (\mathcal{O}_{0:n}^1, \ldots, \mathcal{O}_{0:n}^K).$$

Let $\pi_0(\theta; x)$ (resp. $p(\theta; x, x')$) be the initial distribution (resp. the transition probabilities) associated with parameter θ and $\mathbb{P}_\theta$ the probability distribution of (X_i) on the canonical space and $\mathbb{E}_\theta$ the expectation w.r.t. $\mathbb{P}_\theta$.

3.2. *Likelihood.* Computing the likelihood of a Markov process with transition probabilities $p(\theta; x, x')$ is classical : for population k, it has for expression, using (2.14),

$$(3.4) \quad L(\theta; \mathcal{O}_{0:n}^k) = \pi_0(\theta; x_0^k)(\prod_{i=1}^n p(\theta; x_{i-1}^k, x_i^k))p_1(s_{n+1}^k/x_n^k)p_2(t_{n+1}^k/x_n^k).$$

Joining the observations from the K independent populations, we obtain using Proposition 2.1, (2.10) - (2.13)

$$(3.5) \quad l(\theta; \mathcal{O}_{0:n}(K)) = l_K(\theta) = \sum_{k=1}^K \log L(\theta; \mathcal{O}_{0:n}^k).$$

Reordering the terms of (3.4) and (3.5) according to the parameters yields,

$$(3.6) \quad l(\theta; \mathcal{O}_{0:n}(K)) = \sum_{j=0}^5 l_K^j(\theta).$$

The first term deals with the initial distribution of (S_0, T_0)

$$(3.7) \quad l_K^0(\theta) = \sum_{k=1}^K \log \nu(\theta^1; s_0^k, t_0^k) = l_K^0(\theta^1).$$

Gather in the second term the transition from seeds S, T to rosettes R.

$$(3.8) \quad l_K^1(\theta) = \sum_{k=1}^K \sum_{i=0}^n \log(\mathcal{B}(s_i^k; b) \star \mathcal{B}(t_i^k; b'))(r_i^k) = l_K^1(b, b').$$

The next two terms contain the transitions from R to V and V to F, they write

$$(3.9) \quad l_K^2(\theta) = \sum_{k=1}^K \sum_{i=0}^n v_i^k \, \log c \, + (r_i^k - v_i^k) \, \log(1-c) + C_2(\mathcal{O}_{0:n}(K)) = l_K^2(c),$$

$$(3.10) \quad l_K^3(\theta) = \sum_{k=1}^K \sum_{i=0}^n f_i^k \log d + (v_i^k - f_i^k)\log(1-d) + C_3(\mathcal{O}_{0:n}(K)) = l_K^3(d),$$

where the two terms $C_2(\mathcal{O}_{0:n}(K))$ and $C_3(\mathcal{O}_{0:n}(K))$ only depend on the observations. Set in the next term the transition from F to T,

$$(3.11) \quad l_K^4(\theta) = \sum_{k=1}^K \sum_{i=0}^n \log \left(G(\theta^2; .)^{\star f_i^k} \star \mu(\theta^3; .)\right)(t_{i+1}^k) = l_K^4(\theta^2, \theta^3)$$

The last term concerns the seeds S :

$$(3.12) \qquad l_K^5(\theta) = \sum_{k=1}^{K} \sum_{i=0}^{n} \log(\frac{(\mathcal{M}(s_i^k; a, b) \star \mathcal{M}(t_i^k; a', b'))(s_{i+1}^k, r_i^k)}{(\mathcal{B}(s_i^k; b) \star \mathcal{B}(t_i^k; b'))(r_i^k)})$$

Joining the two terms containing a, a', b, b' yields

$$(3.13) \qquad l_K^6(\theta) = l_K^1(\theta) + l_K^5(\theta) = l_K^6(a, b, a', b').$$

The terms $l_k^0(\theta), l_K^2(\theta), l_K^3(\theta), l_K^4(\theta)$ and $l_K^6(\theta)$ depend on disjoint sets for the parameters. Hence, maximizing the loglikelihood can be performed maximizing separately these five terms.

REMARK 3.1. *Usually, statistical inference for Stochastic Processes is investigated in the asymptotic framework small K (mostly $K = 1$) and large n (leading to asymptotics results $n \to +\infty$). Here, we have that n is small and K large (e.g. magnitude 300). This situation often occurs in Ecology (de Valpine, 2004).*

3.3. *Study of maximum likelihood and other estimators.* This is a K sample of i.i.d. random variables, each variable being a part of a branching process path. We have to use simultaneously the repetitions and the Markov structure to estimate the parameters. So, deriving the properties of the statistical model is not standard. The various terms $l_K^i(\theta)$ of the loglikelihood are associated with different parametric inference problems.

The first term $l_K^0(\theta)$ deals with the estimation of θ^1 based on a sample of K i.i.d. random variables with distribution $\nu(\theta^1; .)$ on $\mathbb{N}^2$, which is standard. The terms $l_K^2(\theta) = l_K^2(c)$ and $l_K^3(\theta) = l_K^3(d)$ are related to the estimation of parameters c, d. Denote by $\hat{c}_K, \hat{d}_K$ the maximum likelihood estimators (MLE) obtained maximizing $l_K^2(c)$ and $l_K^3(d)$. They depend on the successive observations (r_i^k, v_i^k) (resp. (v_i^k, f_i^k)) for $\{i = 0, \ldots, n; k = 1, \ldots, K\}$ and are explicit (see Appendix A.2).

Parameters (a, b, a', b') are only present in $l_K^6(\theta)$ defined in (3.13). Maximum likelihood estimators for (a, b, a', b') can be defined maximizing $l_K^6(\theta)$. To prove identifiability and consistency, we have rather consider here conditional least squares (CLS) estimators. Conditionally on (S_i, T_i), the marginal distribution of S_{i+1} (resp. R_i) is the sum of the two independent distributions, $\mathcal{B}(S_i, a)$ and $\mathcal{B}(T_i, a')$ (resp. $\mathcal{B}(S_i, b)$ and $\mathcal{B}(T_i, b')$). Therefore, we can define the CLS estimators $(\hat{a}_K, \hat{a}_K')$ and $(\hat{b}_K, \hat{b}_K')$ minimizing the Conditional

Least Squares:

$$(3.14) \qquad J_K^1(a, a') = \sum_{k=1}^{K} \sum_{i=0}^{n} (s_{i+1}^k - a s_i^k - a' t_i^k)^2 \; ;$$

$$(3.15) \qquad J_K^2(b, b') = \sum_{k=1}^{K} \sum_{i=0}^{n} (r_i^k - b s_i^k - b' t_i^k)^2.$$

The remaining term is $l_K^4(\theta^2, \theta^3) = l_K^4(m, u)$ since we are concerned by the estimation of m and u. This is the only part of the likelihood associated with the branching mechanism. This likelihood containing the convolution product $G^{\star f} \star \mu$ is untractable, and methods based on conditional least squares or weighted conditional least squares are used for branching processes (Hall and Heyde, 1980; Guttorp, 1991; Wei and Winnicki, 1990) leading to moment estimations of G and μ. Noting that $E(T_{i+1}/F_i) = m F_i + u$, we can just consider for the estimation of m and u, the CLS process,

$$(3.16) \qquad J_K^4(m, u) = \sum_{k=1}^{K} \sum_{i=0}^{n} (t_{i+1}^k - m f_i^k - u)^2.$$

Let $(\hat{m}_K, \hat{u}_K)$ be the CLS estimators minimizing (3.16). All the above estimators are explicitly defined in Appendix A.2 and we can state :

PROPOSITION 3.1. *Assume (A1), (A2), (A3) and (A4). Then, under* $\mathbb{P}_{\theta_0}$, *all the parameters* c, d, a, b, a', b', m, u *are identifiable and, as* $K \to \infty$,
(i) $(\hat{c}_K, \hat{d}_K, \hat{a}_K, \hat{b}_K, \hat{a}'_K, \hat{b}'_K, \hat{m}_K, \hat{u}_K)$ *are consistent and asymptotically Gaussian at rate* $\sqrt{K}$;
(ii) $\hat{c}_K$, $\hat{d}_K$, $(\hat{a}_K, \hat{b}_K, \hat{a}'_K, \hat{b}'_K)$, $(\hat{m}_K, \hat{u}_K)$ *are asymptotically independent, with explicit covariance matrix given in Appendix A.2.*

Let us stress that, before studying in detail this inference problem, it was difficult to assert the classical properties stated in Proposition 3.1. Adding immigration could lead to non identifiability or estimating problems for m and u. Moreover, maximum likelihood estimators, for multitype branching processes, are based on the observations of $G_{i,j}(k)$, i.e. offspring of type j from type i parents (see Guttorp, 1991; Gonzalez et al., 2008; Maaouia and Touati, 2005). We did not require this information for the inference and just used the counts of individuals of each type in successive generations. Hence, getting identifiability and consistency for the parameters is the only difficulty here. Other properties are classical but requires using the exact structure of the data provided the regularity of the statistical model.

REMARK 3.2. *Estimating additional moments of $G(.)$ and $\mu(.)$ can be performed similarly using other functionals than Conditional Least Squares (see Winnicki, 1991 for the variance estimation of $G(.)$ and $\mu(.)$).*

The proof is given in Appendix A.2.

4. Incomplete model study in the Poisson case. The set-up is now different: only R_i, V_i, F_i are observed while S_i and T_i are unobserved. Clearly, algorithms simulating the missing data given the model and the parameters at each step can be used to get estimation. This approach is complementary to our concern that aims at understanding which mechanisms can be estimated. For this, we have to study the process (R_i, V_i, F_i) for $i = 0, \ldots, n$. It is a discrete time stochastic process, which is no longer Markov : the distribution of $(R_{i+1}, V_{i+1}, F_{i+1})$ given the past now depends on the whole past and not only on (R_i, V_i, F_i). This appears explicitely later on. This is similar to problems encountered when studying Hidden Markov Models (Genon-Catalot et al., 2003; Genon-Catalot and Larédo, 2006; Cappé et al., 2005). For a first approach, we restrict our attention to a very informative example, the case of Poisson distributions, which leads to explicit computations.

4.1. *Probabilistic properties in the Poisson case.* Let us specify all the distributions appearing in the populations dynamics.

ASSUMPTION 5. *The offspring distribution $G(.)$ is Poisson $\mathcal{P}(m)$, the immigration distribution $\mu(.)$ is Poisson $\mathcal{P}(u)$.*

ASSUMPTION 6. *The variables S_0 and T_0 are independent and distributed according to Poisson distributions: $S_0 \sim \mathcal{P}(\sigma)$ and $T_0 \sim \mathcal{P}(\tau)$.*

For Poisson distributions, we denote $\mathcal{P}(\lambda)(k) = \mathbb{P}(X = k)$. Recall a property of Multinomial and Poisson distributions.

LEMMA 4.1. *Assume that N is a random variable distributed according to a Poisson distribution $\mathcal{P}(\lambda)$ and that $X = (X_1, X_2, \ldots, X_l)$ is a l-dimensional random variable such that the conditional distribution of X given N is a Multinomial distribution $\mathcal{M}(N; a_1, \ldots, a_l)$ with $\sum_{i=1}^{l} a_i = 1$. Then, the random variables $\{X_i, i = 1, \ldots, l\}$ are independent and verify $X_i \sim \mathcal{P}(a_i\lambda)$.*

First consider one population and omit the index k in what follows. Set

(4.1) $$Y_i = (R_i, V_i, F_i) \quad \text{and} \quad \mathcal{G}_i = \sigma(Y_j, \ j = 0, \ldots, i).$$

Clearly, $\mathcal{G}_i$ is the information available up to time i. To state the main result of this section, define the three sequences of $\mathcal{G}_{i-1}$ measurable random variables:

$$(4.2) \qquad \Gamma_0 = \sigma \; ; \; \Gamma'_0 = \tau \; ;$$

$$(4.3) \qquad \text{for } i \geq 1, \;\; \Gamma_i = a\Gamma_{i-1} + a'\Gamma'_{i-1} \; ; \;\; \Gamma'_i = mF_{i-1} + u.$$

$$(4.4) \qquad \text{for } i \geq 0, \;\; \Lambda_i = b\Gamma_i + b'\Gamma'_i.$$

Then, the following holds :

THEOREM 4.1. *Under Assumptions (A1)-(A6), the initial distribution $\tilde{\pi}_0(y)$ of (Y_i), and the conditional distribution $\mathcal{L}(Y_{i+1}/\mathcal{G}_i)$ satisfy, using (4.1) and definitions (2.13),(4.4), for $y = (r, v, f) \in \mathbb{N}^3$*

$$(4.5) \qquad \mathbb{P}(Y_0 = (r_0, v_0, f_0)) = \tilde{\pi}_0(y_0) = \mathcal{P}(\Lambda_0)(r_0)p_4(v_0/r_0)p_5(f_0/v_0),$$
$$(4.6) \qquad \mathbb{P}(Y_{i+1} = (r, v, f)/\mathcal{G}_i) = \mathcal{P}(\Lambda_{i+1})(r)p_4(v/r)p_5(f/v).$$

The explicit dependence of R_i on the whole past $(Y_0, \ldots, Y_{i-1})$ appears more simply with the following expression for Λ_i,

$$(4.7) \quad \Lambda_0 = b\sigma + b'\tau = c_0(\theta) \; ;$$
$$(4.8) \quad \Lambda_1 = b'mF_0 + c_1(\theta) \;\; \text{with } c_1(\theta) = ab\sigma + a'b\tau + b'u,$$
$$(4.9) \quad \Lambda_i = b'mF_{i-1} + a'bm(F_{i-2} + aF_{i-3} + \cdots + a^{i-2}F_0) + c_i(\theta), \;\; \text{with}$$
$$(4.10) \qquad c_i(\theta) = a^i b\sigma + a^{i-1}a'b\tau + a'bu\,\frac{1 - a^{i-1}}{1 - a} + b'u \;\; \text{for } i \geq 2.$$

The result of Theorem 4.1 is a consequence of the proposition stated below.

PROPOSITION 4.1. *Under Assumptions (A1)-(A6), the random variables S_{i+1}, T_{i+1} are conditionally independent given $\mathcal{G}_i$, and their conditional distributions satisfy, using the random variables Γ_i and Γ'_i defined in (4.2), (4.3),*

$$(4.11) \qquad \mathcal{L}(S_{i+1}/\mathcal{G}_i) \sim \mathcal{P}(\Gamma_{i+1}) \;\; \text{and} \;\; \mathcal{L}(T_{i+1}/\mathcal{G}_i) \sim \mathcal{P}(\Gamma'_{i+1}).$$

Let us prove Theorem 4.1 assuming Proposition 4.1. We just have to check the expression of the conditional distribution of R_{i+1}. By (2.5), we have $R_{i+1} = R'_{i+1} + R''_{i+1}$. Using Proposition 2.1 and (2.12), the distribution

of R_{i+1} conditionnally on (S_{i+1}, T_{i+1}) is equal to $p_3(r/S_{i+1}, T_{i+1})$. Applying Proposition 4.1, S_{i+1} and T_{i+1} are conditionally independent given $\mathcal{G}_i$ and distributed according to two independent Poisson distributions. Hence, another application of Lemma 4.1 yields that the conditional distribution of R_{i+1} given $\mathcal{G}_i$ is $\mathcal{P}(b\Gamma_{i+1} + b'\Gamma'_{i+1}) = \mathcal{P}(\Lambda_{i+1})$, which is (4.4).

The proof of Proposition 4.1 is given in Appendix A.3.

4.2. *Likelihood of the incomplete observations.* The inference is now based on the observations of (Y_i) recorded up to time n for K independent populations. Denote by $Y_i^k = (R_i^k, V_i^k, F_i^k)$ the process describing its dynamics in population k and $y_i^k = (r_i^k, v_i^k, f_i^k)$ the observations at time i. We set,

$$(4.12) \qquad Y_{0:n}^k = (Y_0^k, \ldots, Y_n^k), \quad \text{and} \quad Y_{0:n}(K) = (Y_{0:n}^1, \ldots, Y_{0:n}^K).$$

Observations up to time n are denoted

$$(4.13) \qquad \tilde{O}_{0:n}^k = (y_0^k, y_1^k, \ldots, y_n^k) \quad \text{and} \quad \tilde{O}_{0:n}(K) = (\tilde{O}_{0:n}^1, \ldots, \tilde{O}_{0:n}^K).$$

Let us first compute the likelihood for one population, population k. Successive conditionnings yield

$$(4.14) \qquad L(\theta; \tilde{O}_{0:n}^k) = P_\theta(Y_0^k = y_0^k) \prod_{i=1}^n P_\theta(Y_i^k = y_i^k / Y_{0:i-1}^k = y_{0:i-1}^k).$$

Contrary to the previous section, each term of this product depends on i and on the observations up to time $i - 1$. Theorem 4.1 gives the expression of these conditional probabilities.

Since the random variables Λ_i now depend on θ and on the past, we define $\Lambda_i(\theta) = \Lambda_i(\theta; Y_{0:i-1})$, and for population k,

$$(4.15) \qquad \Lambda_i^k(\theta) = \Lambda_i(\theta; Y_{0:i-1}^k), \quad \lambda_i^k(\theta) = \Lambda_i(\theta; y_{0:i-1}^k).$$

$$(4.16) \qquad P_\theta(Y_i^k = y_i^k / y_{0:i-1}^k) = \mathcal{P}(\lambda_i^k(\theta))(r_i^k) \, p_3(c; v_i^k / r_i^k) \, p_4(d; f_i^k / v_i^k).$$

Joining the observations in the K populations, the likelihood writes, using notations (4.12), (4.13), (4.16),

$$(4.17) \qquad L(\theta; \tilde{O}_{0:n}(K)) = \prod_{k=1}^K L(\theta; \tilde{O}_{0:n}^k).$$

The log-likelihood splits into three terms,

$$(4.18) \qquad \tilde{l}(\theta, \tilde{O}_{0:n}(K)) = \log L(\theta; \tilde{O}_{0:n}(K)) = \sum_{i=1}^3 \tilde{l}_K^i(\theta, \tilde{O}_{0:n}(K)),$$

where, using Assumptions (A5), (A6) and Theorem 4.1,

$$(4.19) \qquad \tilde{l}_K^1(\theta) = \tilde{l}_K^1(\theta, \tilde{O}_{0:n}(K)) = \sum_{k=1}^{K} \sum_{i=0}^{n} \log \mathcal{P}(\lambda_i^k)(r_i^k),$$

$$(4.20) \qquad \tilde{l}_K^2(\theta, \tilde{O}_{0:n}(K)) = l_K^2(c, O_{0:n}) \; ; \; \tilde{l}_K^3(\theta, \tilde{O}_{0:n}(K)) = l_K^3(d, O_{0:n}).$$

Hence, estimating parameters c, d is exactly the same as in the previous section; their inference is omitted in the sequel.

4.3. *Parametric inference.* It remains to study the estimation of the parameter $\theta = (\sigma, \tau, a, b, a', b', u, m)$. As before, let θ_0 be true value of the parameter.

We first have to investigate which parameters are identifiable when only these incomplete observations are available. By identifiability, we mean here identifiability of a statistical model $\mathcal{M} = (\mathbb{P}_\theta, \theta \in \Theta)$:

$$\forall \theta, \; \theta' \in \Theta, \{\mathbb{P}_\theta = \mathbb{P}_{\theta'}\} \Rightarrow \{\theta = \theta'\}.$$

Recall that n denotes the time index of a plant lifecycle and that we consider populations recorded up to time n ($n = 0$ corresponding here to the observation of one complete cycle). Using now the definitions of the terms $c_i(\theta)$ given in (4.7), the following holds :

THEOREM 4.2. *Assume (A1), (A2), (A3), (A4), (A5), (A6). Then,*

1. *if $n = 0$, only $c_0(\theta) = b\sigma + b'\tau$ is identifiable;*
2. *if $n = 1$, $(b'm, c_0(\theta), c_1(\theta))$ is identifiable;*
3. *if $n = 2$, $(\frac{a'b}{b'}, b'm, c_0(\theta), c_1(\theta), c_2(\theta))$, is identifiable;*
4. *if $n \geq 3$, and $a \neq \frac{a'b}{b'}$, then $\phi = (a, \frac{a'b}{b'}, b'm, b'u, b\sigma, b'\tau)$ is identifiable; if $n \geq 3$ and $a = \frac{a'b}{b'}$, then $(a = \frac{a'b}{b'}, b'm, b'u, b\sigma + b'\tau)$.*

Note that identifiability of additional parameters cannot be gained increasing n beyond 3. Larger values of n result in improving the asymptotic variance for the estimation of ϕ.

REMARK 4.1. *Stating the above theorem for the first values of n is unusual. However, in the field survey of feral oilseed rape populations, observations unfortunately had been collected up to $n = 2$, leading to the unability of estimating a, the annual survival rate in the seed bank, which is a parameter of much concern in Ecology.*

Assume now that $n \geq 3$, $a \neq \frac{a'b}{b'}$ and let us study the inference of the identifiable parameters. Let us denote by $\phi = (a, \frac{a'b}{b'}, b'm, b'u, b\sigma, b'\tau)$ (resp. ϕ_0) an arbitrary (resp. the true value) of the paramete, by Φ the parameter set and asssume

ASSUMPTION 7. Φ is a compact set of $\mathbb{R}_+^{*\,5} \times (0,1)$; and $\phi_0 \in \overset{\circ}{\Phi}$.

Using (4.19), define the maximum likelihood estimator $\hat{\phi}_K$ as a solution of

$$(4.21) \qquad \tilde{l}_K^1(\hat{\phi}_K) = \sup\{\tilde{l}_K^1(\phi) \,; \phi \in \Phi\}.$$

Under Assumption (A7), the function $\phi \to \sum_0^n (\Lambda_i(\phi; Y_{0:i-1}) - \log \Lambda_i(\phi, Y_{0:i-1}))$ is a.s. $\mathbb{P}_\phi$ twice differentiable on Φ, and we can define, for $\phi = (\phi_1, \ldots, \phi_6)$, the 6×6 matrix,

$$I(\phi)_{p,q} = \sum_{i=0}^{n} \mathbb{E}_{\theta_0}\left(\frac{1}{\Lambda_i(\phi)} \frac{\partial \Lambda_i(\phi)}{\partial \phi_p} \frac{\partial \Lambda_i(\phi)}{\partial \phi_q}\right) \text{ with } 1 \leq p, q \leq 6.$$

PROPOSITION 4.2. Assume (A1)-(A3), (A5), (A6) and (A7). Then $\hat{\phi}_K$ is strongly consistent. If moreover the matrix $I(\phi_0)$ is invertible, then

$$\sqrt{K}(\hat{\phi}_K - \phi_0) \xrightarrow{\mathcal{D}} \mathcal{N}(0, I(\phi_0)^{-1}) \text{ under } \mathbb{P}_{\phi_0} \text{ as } K \to +\infty.$$

REMARK 4.2. The matrix $I(\phi_0)$ that appears in the asymptotic variance of $\hat{\phi}_K$ can be estimated, using the explicit expressions for the derivatives of $\Lambda_i(\phi)$, by the empirical estimate for $1 \leq p, q \leq 6$,

$$\hat{I}_{p,q} = \frac{1}{K} \sum_{k=1}^{K} \sum_{i=0}^{n} \frac{1}{\lambda_i^k(\hat{\phi}_K)} \frac{\partial \lambda_i^k(\hat{\phi}_K)}{\partial \phi_p} \frac{\partial \lambda_i^k(\hat{\phi}_K)}{\partial \phi_q}.$$

The proofs of Theorem 4.2 and Proposition 4.2 are given in Appendix A.4.A.5.

5. Simulation study. In the case of incomplete observations, the estimators strongly rely on the assumption that both the offspring and the immigration distributions are Poisson distributions. We investigate in this section how the estimation behaves when these distributions are no longer Poisson.

5.1. *Methods.* We considered in these simulations that the offspring distribution $G(.)$ or the immigration distribution $\mu(.)$ are more dispersed than Poisson distributions. Two series of simulations were performed: the first one concerns deviations to a Poisson distribution of the offspring $G(.)$, and the other one for the immigration $\mu(.)$. For this, we used Negative Binomial distributions for $G(.)$ (resp. $\mu(.)$) with mean m (resp. u) and increasing values of the variance variance δ^2 (resp. ρ^2); the (variance /mean) ratio ranged from 2 to 1000. For each given set of parameters, we performed $M = 100$ repetitions including $K = 300$ populations. These populations dynamics were run during $n = 4$ years to get the observations for rosettes R_0 to R_4 and mature plants (F_0 to F_3), using biologically plausible values for demographic parameters. Indeed, for a, b, a', b', we used the parameters given in Claessen et al., 2005a,b :

(5.1) $$a = 0.15 \ , \ a' = 0.006 \ , \ b = 0.5 \ , \ b' = 0.5.$$

We also had to fix some values for $c, d, m = \mathbb{E}(G), u = \mathbb{E}(\mu)$. We used values estimated in Garnier et al., 2008 :

(5.2) $$c = 0.21 \ , \ d = 0.01 \ , \ m = 13 \ , \ u = 80.$$

The value $m = 13$ corresponds to the mean fecundity of plants mown twice. (Colbach et al., 2001). The value $u = 80$ corresponds to average immigration when there is no cultivated field in the neighbourhood (Garnier et al., 2008). Finally, for the initial distributions of S_0, T_0, we assumed

(5.3) $$S_0 \sim \mathcal{P}(\sigma), T_0 \sim \mathcal{P}(\tau) \text{ with } \sigma = \tau = 50.$$

For each value of (variance/mean) ratio, the inference of the identifiable parameters (see Theorem 4.2) in the case $n \geq 3$. was computed on each of the $M = 100$ repetitions of the $K = 300$ populations trajectories. Mean and standard deviation of these estimates were then empirically estimated from the $M = 100$ values obtained. Simulations and statistical analyses were performed using R-8-0 (R Development Core Team, 2005).

5.2. *Results.* Some technical difficulties occured when we tried to estimate jointly the six identifiable parameters $(a, \frac{a'b}{b'}, b'm, b'u, b\sigma, b'\tau)$: the algorithm we used often did not converge because of the non-linearity in the statistical model. Since we were mainly interested in the estimations for $G(.)$ and $\mu(.)$, we assumed that the quantities a and $\frac{a'b}{b'}$ were known. With this simplification, we just had to deal with a linear statistical model; we restricted our attention to the estimation of the parameters $(b'm, b'u, b\sigma, b'\tau)$.

δ^2/m	$b'm = 6.5$		$b'u = 40$		$b'\sigma = 25$		$b'\tau = 25$	
	est.	sd	est.	sd	est.	sd	est.	sd
2	6.44	0.77	40.02	0.21	25.19	3.28	24.88	3.27
5	6.46	0.61	40.05	0.24	24.46	3.34	25.52	3.36
10	6.65	0.84	39.96	0.23	25.18	2.93	24.84	2.95
50	6.78	1.40	39.97	0.27	25.10	3.73	24.88	3.75
100	6.56	1.89	39.99	0.25	25.24	3.63	24.73	3.64
500	6.26	3.30	40.01	0.32	24.45	6.66	25.50	6.69
1000	6.41	5.41	40.003	0.38	25.54	7.75	24.37	7.78

TABLE 1

Mean (est.) and standard deviation (sd) of estimators when the offspring $G(.)$ is a binomial distribution with mean m and variance δ^2 instead of a $\mathcal{P}(m)$ distribution. Immigration $\mu \sim \mathcal{P}(u)$; $a = 0.16, a' = 0.006, b = b' = 0.5, m = 13, u = 80, \sigma = \tau = 50$.

Value ρ^2/u	$b'm = 6.5$		$b'u = 40$		$b'\sigma = 25$		$b'\tau = 25$	
	est.	sd	est.	sd	est.	sd	est.	sd
2	6.51	0.77	40.01	0.24	24.61	3.81	25.36	3.76
5	6.45	0.98	40.01	0.37	24.33	5.19	25.67	5.28
10	6.61	1.47	40.03	0.52	24.78	7.25	25.17	7.25
50	6.59	2.63	39.93	0.87	25.60	14.89	24.46	14.93
100	7.42	3.43	39.95	1.38	28.60	22.42	21.39	22.40
500	7.61	7.26	40.05	3.13	21.96	39.31	28.06	39.31
1000	6.48	7.63	39.62	4.23	21.77	64.75	28.23	64.76

TABLE 2

Mean (est.) and standard deviation (sd) of estimators when the immigration $\mu(.)$ is a negative binomial distribution with mean u and variance ρ^2 instead of a $\mathcal{P}(u)$ ditribution. Offspring $G(.) \sim \mathcal{P}(m)$; $a = 0.16, a' = 0.006, b = b' = 0.5, m = 13, u = 80, \sigma = \tau = 50$.

The results are given in the two tables below.

Concerning deviations of $G(.)$ from the distribution $\mathcal{P}(m)$, we obtained that the estimation procedure performed very well, even for large deviations from the Poisson case : the bias remained less than 5% for all the four estimated quantities with a variance/ratio up to 1000. (see Table 5.2) When immigration was assumed to follow a Negative Binomial distribution, the estimation procedure performed quite well for values of (variance/mean) ratios up to 50 : the biases remained less than 10% for all four identifiable quantities. For larger values of variance/mean ratio, the bias could 17% of parameter value approximately.

APPENDIX A: APPENDIX SECTION

A.1. Proof of Proposition 2.1 . Consider first the initial distribution $\pi_0(x)$. Successive conditionings yield

$$\pi_0(x) = \mathbb{P}(S_0 = s, T_0 = t)\mathbb{P}(R_0 = r/s, t)\mathbb{P}(V_0 = v/s, t, r)\mathbb{P}(F_0 = f/s, t, r, v).$$

The first distribution is $\nu(s, t)$. Using definitions (2.4), (2.13), the last two conditional distributions are equal to $p_4(v/r)$ and $p_5(f/v)$. The remaining distribution in $\pi_0(x)$ is $\mathcal{L}(R_0/S_0, T_0)$. Using (A2) and (2.5), $R_0 = R_0' + R_0''$, where R_0' and R_0'' are independent. Now, the distribution of (S_1', R_0') (resp. (S_1'', R_0'')) is the Multinomial distribution $\mathcal{M}(S_0; a, b)$ (resp. $\mathcal{M}(T_0; a', b')$), leading to Binomial distributions $\mathcal{B}(S_0; b)$ for the marginal distribution of R_0' (resp. $\mathcal{B}(T_0; b')$ for R_0''). By (A2), these two distributions are independent conditionally on (S_0, T_0) which yields (2.12) and the expression of $\pi_0(x)$. Let us now study two successive generations. Using notations (2.7), the conditional distribution of X_{i+1} given $\mathcal{F}_i$ can be expressed for $x = (s, t, r, v, f), x' = (s', t', r', v', f')$,

$$\mathbb{P}(X_{i+1} = x'/\mathcal{F}_i) = \mathbb{P}(S_{i+1} = s'/\mathcal{F}_i) \times \mathbb{P}(T_{i+1} = t'/s'; \mathcal{F}_i) \times$$
$$\mathbb{P}(R_{i+1} = r'/s', t'; \mathcal{F}_i) \times \mathbb{P}(V_{i+1} = v'/s', t', r'; \mathcal{F}_i) \times \mathbb{P}(F_{i+1} = f'/s', t', r', v'; \mathcal{F}_i).$$

The last three conditional distributions have already been computed for getting $\pi_0(x)$. They are respectively equal to $p_3(r'/s', t')$, $p_4(v'/r')$ and $p_5(f'/v')$ defined in (2.12), (2.13). Let us compute $\mathbb{P}(T_{i+1} = t'/s'; \mathcal{F}_i)$. Seeds on the ground at cycle $(i + 1)$ come from two sources, offspring of mature plants F_i and seed immigration during cycle $i + 1$, $T_{i+1} = T_{i+1}' + I_{i+1}$. Using (A1), the distribution of T_{i+1}' is $G^{\star F_i}$. By (A3), I_{i+1} is independent of S_{i+1}, T_{i+1}' and $\mathcal{F}_i$, so that $\mathbb{P}(T_{i+1} = t'/s'; \mathcal{F}_i) = (G^{\star f} \star \mu)(t') = p_2(t'/f)$. The last distribution is $\mathbb{P}(S_{i+1} = s'/\mathcal{F}_i)$. Using (A2) and (2.5) yields that $S_{i+1} = S_{i+1}' + S_{i+1}''$

where conditionally on (S_i, T_i), (S'_{i+1}, R'_i) and (S''_{i+1}, R''_i) are distributed according to two independent Multinomial distributions $\mathcal{M}(S_i; a, b)$ and $\mathcal{M}(T_i; a', b')$. Hence, the marginal distribution of S_{i+1} is,

$$\mathbb{P}(S_{i+1} = s'/\mathcal{F}_i) = \frac{(\mathcal{M}(S_i; a, b) \star \mathcal{M}(T_i; a', b'))(s', r)}{(B(S_i, b) \star B(T_i, b'))(r)} = p_1(s'/s, t, r).$$

Joining these results yields that (X_i) is a time homogeneous Markov chain with state space $\mathbb{N}^5$.

A.2. Proof of Proposition 3.1. We use in the sequel the Kullback-Leibler divergence $\mathcal{K}(P, Q)$ of distribution Q w.r.t. P. Recall its definition,

$$(A.1) \qquad \mathcal{K}(P, Q) = -\int \log \frac{dQ}{dP} \, dP = -\mathbb{E}_{\mathbb{P}}(\frac{dQ}{dP}) \text{ if } Q \ll P \,;$$

$$(A.2) \qquad \mathcal{K}(P, Q) = +\infty \quad \text{otherwise.}$$

This quantity is non-negative and equal to 0 if and only if $Q = P$ P a.s.

Let us first consider $l_K^2(c)$ and $l_K^3(d)$ defined in (3.9), (3.10). The maximum likelihood estimators are

$$(A.3) \qquad \hat{c}_K = \frac{\sum_{k=1}^{K} \sum_{i=0}^{n} v_i^k}{\sum_{k=1}^{K} \sum_{i=0}^{n} r_i^k} \;;\; \hat{d}_K = \frac{\sum_{k=1}^{K} \sum_{i=0}^{n} f_i^k}{\sum_{k=1}^{K} \sum_{i=0}^{n} v_i^k}.$$

Since the K populations are independent, applying the strong law of large numbers yields the strong consistency of $\hat{c}_K$ and $\hat{d}_K$. The random variables $Z_i^k = \sum_{i=0}^{n}(V_i^k - c_0 R_i^k)$ are i.i.d. centered with variance $c_0(1 - c_0)\mathbb{E}_{\theta_0}(\sum_{i=0}^{n} R_i)$, so that the Central Limit Theorem yields that,

$$(A.4) \quad \text{as } K \to +\infty \,,\; \sqrt{K}(\hat{c}_K - c_0) \xrightarrow{\mathcal{D}} \mathcal{N}(0, \frac{c_0(1 - c_0)}{\mathbb{E}_{\theta_0}(\sum_{i=0}^{n} R_i)}) \text{ under } \mathbb{P}_{\theta_0}.$$

The proof concerning $\hat{d}_k$ is similar : $\hat{d}_K \to d_0$ $\mathbb{P}_{\theta_0}$ a.s. and

$$(A.5) \qquad\qquad \sqrt{K}(\hat{d}_K - d_0) \xrightarrow{\mathcal{D}} \mathcal{N}(0, \frac{d_0(1 - d_0)}{\mathbb{E}_{\theta_0}(\sum_{i=0}^{n} V_i)}).$$

Consider now the estimation of (a, b, a', b'). Let us first check identifiability. Applying the strong law of large numbers to $l_K^6(\theta; \mathcal{O}_{0:n})$ defined in (3.13), we get that, under $\mathbb{P}_{\theta_0}$, as $K \to +\infty$,

$$\frac{1}{K} l_K^6(\theta; X_{0:n}(K)) \to \mathbb{E}_{\theta_0} \sum_{i=0}^{n} \log(\mathcal{M}(S_i; a, b) \star \mathcal{M}(T_i; a', b'))(S_{i+1}, R_i).$$

We can express the limit above using the Kullback-Leibler divergence defined in (A.1),

$$C(\theta_0) - \mathbb{E}_{\theta_0} \sum_{i=0}^{n} \mathcal{K}(\mathcal{M}(S_i; a_0, b_0) \star \mathcal{M}(T_i; a'_0, b'_0), \mathcal{M}(S_i; a, b) \star \mathcal{M}(T_i; a', b')),$$

where $C(\theta_0)$ is a constant depending only on θ_0, and $\mathcal{K}(P, Q)$ is the Kullback-Leibler divergence of the two random probability distributions. Each term of the sum above is non positive and equal to 0 if and only if the two distributions are identical a.s. under $\mathbb{P}_{\theta_0}$. It is easy to check, using the first and second moments of these two distributions, that this implies $(a, b, a', b') = (a_0, b_0, a'_0, b'_0)$. Hence, (a_0, b_0, a'_0, b'_0) is a strict maximum of the limit above, which leads to the identifiability of these parameters.

Consider now the CLS estimators $(\hat{a}_K, \hat{a}'_K)$ which minimize $J_K^1(a, a')$ defined in (3.14). By the strong law of large numbers, under $\mathbb{P}_{\theta_0}$, as $K \to \infty$,

$$(A.6) \qquad \frac{1}{K} J_K^1(a, a') \to \sum_{i=0}^{n} \mathbb{E}_{\theta_0}(\{(a_0 - a)S_i + (a'_0 - a')T_i\}^2) + A(\theta_0),$$

where $A(\theta_0) = a_0(1 - a_0)E_{\theta_0}(\sum_{i=0}^{n} S_i) + a'_0(1 - a'_0)\mathbb{E}_{\theta_0}(\sum_{i=0}^{n} \mathbb{E}_{\theta_0} T_i)$ only depends on θ_0. Since S_i and T_i are non negative random variables, this limit possesses a strict minimum at $(a, a') = (a_0, a'_0)$.

Denote by $^\tau M$ the transposition of a matrix M and set Z the $(n \times 2)$ matrix with rows equal to (S_i, T_i), $^\tau S$ the vector $(S_0, \ldots, S_n)$, $^\tau \tilde{S}$ the vector $(S_1, \ldots, S_{n+1})$, and $Z^k, S^k, \tilde{S}^k, T^k$ their values for population k. Then $J_K^1(a, a')$ writes,

$$\frac{1}{K} J_K^1(a, a') = \frac{1}{K} \sum_{i=1,\ldots,K} {}^\tau\left(\tilde{S}^k - Z^k \begin{pmatrix} a \\ a' \end{pmatrix}\right)\left(\tilde{S}^k - Z^k \begin{pmatrix} a \\ a' \end{pmatrix}\right)$$

$$(A.7) \qquad \text{so that} \quad \begin{pmatrix} \hat{a}_K \\ \hat{a}'_K \end{pmatrix} = \left(\sum_{k=1}^{K} {}^\tau Z^k\, Z^k\right)^{-1}\left(\sum_{k=1}^{K} {}^\tau Z^k \tilde{S}^k\right).$$

As functions of (a, a'), $\frac{1}{K} J_K^1(a, a')$ and its limit defined in (A.6) are twice continuously differentiable a.s.. The parameter set Θ is compact, so we just have to control the continuity modulus of the process $\frac{1}{K} J_K^1(a, a')$. It is defined, for $\eta > 0$, by

$$w(K, \eta) = sup\{\frac{1}{K}|J_K^1(a_1, a'_1) - J_K^1(a_2, a'_2)| \; ; \; \|(a_1, a'_1) - (a_2, a'_2)\| < \eta\}.$$

By the Cauchy-Schwarz inequality, $w(K, \eta)$ is bounded by

$$\{\frac{1}{K}\sum_{k,i}(2S_{i+1}^k-(a_1+a_2)S_i^k-(a_1'+a_2')T_i^k)^2\}^{1/2}\ \{\frac{1}{K}\sum_{k,i}((a_1-a_2)S_i^k+(a_1'-a_2')T_i^k)^2\}^{1/2}.$$

The first term is a random variable converging $\mathbb{P}_{\theta_0}$-a.s. to a deterministic positive limit, and is thus bounded in probability. The second term is bounded by the r.v. $2\eta(\frac{1}{K}\sum_{k=1}^K\sum_{i=0}^n(S_i^k+T_i^k))^{1/2} \to 2\eta\ (E_{\theta_0}\sum_{i=0}^n(S_i+T_i))^{1/2}\ \mathbb{P}_{\theta_0}$ a.s. Hence, as $K \to +\infty$, limsup $w(K,\eta) \le \phi(\eta)$, where $\phi(\eta) \to 0$ as $\eta \to 0$. This ensures the consistency of $(\hat{a}_K, \hat{a}'_K)$ defined in (A.7) (Dacunha-Castelle and Duflo, 1993; Van der Vaart, 1998).

Consider now the asymptotic normality. The function $(a, a') \to J_K^1(a, a')$ being C^2 a.s., the gradient of $J_K^1(a, a')$ is

$$DJ_K^1(a, a') = -\sum_{k=1}^K {}^\tau Z^k\left(\tilde{S}^k - Z^k\begin{pmatrix} a \\ a' \end{pmatrix}\right).$$

The 2x2 matrix containing the second derivatives of $J_K^1(a, a')$ is

$$\nabla J_K^1(a, a') = 2\sum_{k=1}^K {}^\tau Z^k Z^k.$$

Using now that $(\hat{a}_K, \hat{a}'_K)$ is consistent, a Taylor expansion of DJ_K^1 at (a_0, a_0') yields

$$(A.8) \qquad 0 = \frac{1}{2\sqrt{K}}DJ_K^1(a_0, a_0') + \frac{1}{2\sqrt{K}}\nabla J_K^1\begin{pmatrix} a - a_0 \\ a' - a_0' \end{pmatrix} + o_P(1),$$

where $o_P(1)$ denotes a remainder term that goes to 0 in P_{θ_0}-probability. The r.v. vectors $\left(\tilde{S} - Z^k\begin{pmatrix} a_0 \\ a_0' \end{pmatrix}\right)$ are i.i.d. centered, hence $\frac{1}{2\sqrt{K}}DJ_K^1(a_0, a_0')$ converges in distribution under $\mathbb{P}_{\theta_0}$ to a centered Normal distribution with covariance matrix $E_{\theta_0}({}^\tau ZV(\theta_0)Z)$ where $V(\theta_0)$ is the $(n+1)\times(n+1)$ diagonal matrix with diagonal elements $a_0(1-a_0)S_i + a_0'(1-a_0')T_i$. Moreover, by the strong law of large numbers, $\frac{1}{2K}\nabla J_K^1$ converges a.s. to $E_{\theta_0}({}^\tau ZZ)$. By the Cauchy-Schwarz inequality, this matrix is invertible and using (A.8) yields

$$\sqrt{K}\begin{pmatrix} \hat{a}_K - a_0 \\ \hat{a}'_K - a_0' \end{pmatrix} = \left(E_{\theta_0}({}^\tau ZZ)\right)^{-1}\frac{1}{\sqrt{K}}\sum_{k=1}^K {}^\tau Z^k(S^k - Z^k\begin{pmatrix} a_0 \\ a_0' \end{pmatrix}) + o_P(1).$$

Therefore, setting $\Sigma_1(\theta_0) = (E_{\theta_0}({}^\tau ZZ))^{-1}E_{\theta_0}({}^\tau ZV(\theta_0)Z)(E_{\theta_0}({}^\tau ZZ))^{-1}$,

$$(A.9) \qquad \sqrt{K}\begin{pmatrix} \hat{a}_K - a_0 \\ \hat{a}'_K - a_0' \end{pmatrix} \to^{\mathcal{D}} \mathcal{N}(0, \Sigma_1(\theta_0)).$$

Similarly, setting $^\tau R$ the vector $(R_0, \ldots, R_n)$, define

$$(A.10) \qquad \begin{pmatrix} \hat{b}_K \\ \hat{b}'_K \end{pmatrix} = \Big(\sum_{k=1}^{K} {}^\tau Z^k \, Z^k \Big)^{-1} \Big(\sum_{k=1}^{K} {}^\tau Z^k R^k \Big).$$

The proof concerning the estimation of (b, b') is similar; so we get, setting $W(\theta_0)$ diagonal matrix with diagonal elements $b_0(1 - b_0)S_i + b'_0(1 - b'_0)T_i$ and $\Sigma_2(\theta_0) = (E_{\theta_0}({}^\tau ZZ))^{-1} E_{\theta_0}({}^\tau ZW(\theta_0)Z)(E_{\theta_0}({}^\tau ZZ))^{-1}$,

$$(A.11) \qquad \sqrt{K} \begin{pmatrix} \hat{b}_K - b_0 \\ \hat{b}'_K - b'_0 \end{pmatrix} \to^{\mathcal{D}} \mathcal{N}(0, \Sigma_2(\theta_0)).$$

Finally, let us study the estimation of (m, u) based on the CLS process $J_K^4(m, u)$ defined in (3.16). Under the assumptions (A1), (A3), $G(.)$ and $\mu(.)$ have finite variances δ^2 and ρ^2. Denote by δ_0^2 and ρ_0^2 the variances associated with θ_0. Then, as $K \to \infty$, under $\mathbb{P}_{\theta_0}$,

$$\frac{1}{K} J_K^4(m, u) \to E_{\theta_0} \Big(\sum_{i=0}^{n} (\delta_0^2 F_i + \rho_0^2) \Big) + E_{\theta_0} \Big(\sum_{i=0}^{n} [(m_0 - m)F_i + (u - u_0)]^2 \Big).$$

Clearly, the above functional has a strict minimum at (m_0, u_0), leading to the identifiability of (m, u). Let $^\tau F$ denote the vector $(F_0, \ldots, F_n)$, G the $(n + 1) \times 2$ matrix with rows equal to $(F_i, 1)$, $^\tau \tilde{T}$ the vector $(T_1, \ldots, T_{n+1})$, $F^k, G^k, \tilde{T}^k$ their values in population k. Then, the conditional least square estimator is

$$(A.12) \qquad \begin{pmatrix} \hat{m}_K \\ \hat{u}_K \end{pmatrix} = \Big(\sum_{k=1}^{K} {}^\tau G^k \, G^k \Big)^{-1} \Big(\sum_{k=1}^{K} {}^\tau G^k \tilde{T}^k \Big).$$

Consistency and asymptotic normality are obtained with a proof similar to the one detailed above. Define $W'(\theta_0)$ the diagonal matrix with diagonal elements $\delta_0^2 F_i + \rho_0^2$, and $\Sigma_3(\theta_0) = (E_{\theta_0}({}^\tau GG))^{-1} E_{\theta_0}({}^\tau GW'G)(E_{\theta_0}({}^\tau GG))^{-1}$, then

$$(A.13) \quad \sqrt{K} \begin{pmatrix} \hat{m}_K - m_0 \\ \hat{u}_K - u_0 \end{pmatrix} \to^{\mathcal{D}} \mathcal{N}(0, \Sigma_3(\theta_0)) \quad \text{in distribution under } \mathbb{P}_{\theta_0}.$$

Using now that the likelihood splits into five terms that can be maximized separately leads to the asymptotic independence of the estimators stated in Proposition 3.1.

A.3. Proof of Proposition 4.1. Let us prove by induction Proposition 4.1. By Assumption A6, the property holds for $i = 0 : S_0$ and T_0 are independent and $S_0 \sim \mathcal{P}(\sigma) = \mathcal{P}(\Gamma_0)$, $T_0 \sim \mathcal{P}(\tau) = \mathcal{P}(\Gamma'_0)$. Assume that the property holds for $i \geq 1$, i.e. using definitions (4.1), (4.3), S_i and T_i are independent conditionnally on $\mathcal{G}_{i-1}$ and $\mathcal{L}(S_i/\mathcal{G}_{i-1}) = \mathcal{P}(\Gamma_i)$, $\mathcal{L}(T_i/\mathcal{G}_{i-1}) = \mathcal{P}(\Gamma'_i)$. Using (2.5), $(S_{i+1}, R_i) = (S'_{i+1}, R'_i) + (S''_{i+1}, R''_i)$, where the conditional distribution of (S'_{i+1}, R'_i) (resp.(S''_{i+1}, R''_i)) given (S_i, T_i) is the Multinomial distribution $\mathcal{M}(S_i; a, b)$ (resp. $\mathcal{M}(T_i; a', b')$). Using Assumption (A2), these two distributions are independent conditionally on S_i, T_i and applying now Lemma 4.1 conditionally on $\mathcal{G}_{i-1}$ to $\{S_i \sim \mathcal{P}(\Gamma_i), \mathcal{M}(S_i; a, b)\}$ and to $\{T_i \sim \mathcal{P}(\Gamma'_i), \mathcal{M}(T_i; a', b')\}$ yields that the four variables $S'_{i+1}, R'_i, S''_{i+1}, R'_i$ are independent conditionally on $\mathcal{G}_{i-1}$ and that $S'_{i+1} \sim \mathcal{P}(a\Gamma_i)$, $S''_{i+1} \sim \mathcal{P}(a'\Gamma'_{i-1})$, $R'_i \sim \mathcal{P}(b\Gamma_i)$ and $R''_i \sim \mathcal{P}(b'\Gamma'_i)$. Hence, S_{i+1} and R_i are independent conditionally on $\mathcal{G}_{i-1}$ and $S_{i+1} \sim \mathcal{P}(a\Gamma_i + a'\Gamma'_i)$, and $R_i \sim \mathcal{P}(b\Gamma_i + b'\Gamma'_i)$.

Let us now prove that $\mathbb{E}(S_{i+1}/\mathcal{G}_i) = \mathbb{E}(S_{i+1}/\mathcal{G}_{i-1})$. Let ϕ_1 and ϕ_2 two measurable functions of $(Y_{0:i-1}, S_{i+1})$ and $(Y_{0:i-1}, R_i, V_i, F_i)$ and compute,

$$\mathbb{E}^{\mathcal{G}_{i-1}}(\phi_1(S_{i+1})\phi_2(R_i, V_i, F_i)) = \mathbb{E}^{\mathcal{G}_{i-1}}\left(\int \phi_1(s_{i+1})\phi_2(r_i, v_i, f_i)ds_{i+1}dr_idv_idf_i\right).$$

Using (2.13) and (2.14) , set $\psi_2(r) = \int \phi_2(r, v, f)p_4(v/r)p_5(f/v)dv$, then

$$\mathbb{E}^{\mathcal{G}_{i-1}}(\phi_1(S_{i+1})\phi_2(R_i, V_i, F_i)) = \mathbb{E}^{\mathcal{G}_{i-1}}(\phi_1(S_{i+1})\psi_2(R_i))$$

$$= \mathbb{E}^{\mathcal{G}_{i-1}}(\phi_1(S_{i+1}))\mathbb{E}^{\mathcal{G}_{i-1}}(\psi_2(R_i)) = \mathbb{E}^{\mathcal{G}_{i-1}}(\phi_1(S_{i+1}))\mathbb{E}^{\mathcal{G}_{i-1}}(\phi_2(R_i, V_i, F_i)),$$

since S_{i+1} and R_i are independent conditionally on $\mathcal{G}_{i-1}$. Hence, S_{i+1} and (R_i, V_i, F_i) are independent conditionally on $\mathcal{G}_{i-1}$ and,

$$\mathbb{E}(S_{i+1}/\mathcal{G}_i) = \mathbb{E}(S_{i+1}/\mathcal{G}_{i-1}) = \mathcal{P}(a\Gamma_i + a'\Gamma'_{i+1}).$$

Consider two measurable functions ϕ_3, ϕ_4 of $(Y_{0:i}, S_{i+1})$ and $(Y_{0:i}, T_{i+1})$, then using (2.12), (2.14), Assumption (A2) and the conditional independence given $\mathcal{G}_{i-1}$ of S_{i+1} and (R_i, V_i, F_i) yield,

$$(A.14) \quad \mathbb{E}^{\mathcal{G}_i}(\phi_3(S_{i+1})\phi_4(T_{i+1})) = \mathbb{E}^{\mathcal{G}_i}\left(\int (\phi_3(s_{i+1})\phi_4(t')p_2(t'/F_i)ds_{i+1}dt'\right)$$

$$= \mathbb{E}^{\mathcal{G}_i}(\phi_3(S_{i+1}))\left(\int \phi_4(t')p_2(t/F_i)dt'\right) = \mathbb{E}^{\mathcal{G}_i}(\phi_3(S_{i+1}))\mathbb{E}^{\mathcal{G}_i}(\phi_4(T_{i+1})).$$

Therefore, conditionally on $\mathcal{G}_i$ S_{i+1} and T_{i+1} are independent and using now Assumption (A5), $\mathcal{L}(T_{i+1}/\mathcal{G}_i) = \mathcal{P}(mF_i + u)$. The property holds for $i + 1$ with $\Gamma_{i+1} = a\Gamma_i + a'\Gamma'_i$ and $\Gamma'_{i+1} = mF_i + u$, which are the definitions given in (4.3).

A.4. Proof of Theorem 4.2. The likelihood $l_K^{\tilde{1}}(\theta)$ defined in (4.19) sums up the available information associated with the incomplete model. An application of the strong law of large numbers yields,

$$\frac{1}{K} l_K^{\tilde{1}}(\theta) \to L(\theta_0, \theta) = \mathbb{E}_{\theta_0} \sum_{i=0}^{n} (-\Lambda_i(\theta, Y_{0:i-1}) + R_i \log \Lambda_i(\theta, Y_{0:i-1})).$$

According to Theorem 4.1, conditionally on $\mathcal{G}_{i-1}$, the random variables R_i are Poisson distributions with parameter $\Lambda_i(\theta_0, Y_{0:i-1})$ under $\mathbb{P}_{\theta_0}$. Hence,

$$L(\theta_0, \theta) = \mathbb{E}_{\theta_0} \sum_{i=0}^{n} (-\Lambda_i(\theta, Y_{0:i-1}) + \Lambda_i(\theta_0, Y_{0:i-1}) \log \Lambda_i(\theta, Y_{0:i-1})).$$

Using the explicit form of the Kullback-Leibler divergence between two Poisson distributions, $L(\theta_0, \theta)$ writes,

$$L(\theta_0, \theta) = -\sum_{i=0}^{n} \mathbb{E}_{\theta_0} \{ \mathcal{K}(\mathcal{P}(\Lambda_i(\theta_0, Y_{0:i-1})), \mathcal{P}(\Lambda_i(\theta, Y_{0:i-1}))) \} + C(\theta_0),$$

with $C(\theta_0) = \mathbb{E}_{\theta_0} \sum_{i=0}^{n} (-\Lambda_i(\theta_0, Y_{0:i-1}) + \Lambda_i(\theta_0, Y_{0:i-1}) \log \Lambda_i(\theta_0, Y_{0:i-1}))$ only depends on the observations. The identifiability condition for θ_0 is therefore equivalent to

$$\{L(\theta_0, \theta) = 0\} \Rightarrow \{\theta = \theta_0\}.$$

The Kullback-Leibler divergence of two Poisson distributions $\mathcal{K}(\mathcal{P}(\mu_0), \mathcal{P}(\mu))$ is non negative and equal to 0 if and only if $\mu = \mu_0$. Hence, the limit $L(\theta_0, \theta)$ presents a strict maximum at θ_0 if and only if

$$\Lambda_i(\theta, Y_{0:i-1}) = \Lambda_i(\theta_0, Y_{0:i-1}) \, \mathbb{P}_{\theta_0} - \text{a. s. for } i = 0, \dots, n.$$

Since the $\Lambda_i(\theta, Y_{0:i-1})$ depend on $(F_{i-1}, \dots, F_0)$, the above condition can hold only if the coefficients associated with the random variables F_i in the above expression are equal. The proof below is just elementary algebra based on this property. We obtain, using for the $c_i(\theta)$ the definitions given in (4.7) and (4.15) :

If $n = 0$, $\Lambda_0(\theta) = c_0(\theta)$ is deterministic. Only $c_0(\theta)$ is identifiable: it is the first condition stated in Theorem 4.2.

If $n = 1$, we have, $\Lambda_1(\theta, Y_{0:i-1}) = b'm F_0 + c_1(\theta)$. Since F_0 is random, the two random variables $\Lambda_1(\theta, Y_{0:i-1})$ and $\Lambda_1(\theta_0, Y_{0:i-1})$ are $\mathbb{P}_{\theta_0}$-a.s. equal iff $b'm = b'_0 m_0$ and $c_1(\theta) = c_1(\theta_0)$, which leads to the identifiability of $(b'm, c_0(\theta), c_1(\theta))$.

If $n = 2$, $\Lambda_2(\theta) = b'm \, F_1 + \frac{a'b}{b'} b'm F_0 + c_2(\theta)$, We thus get two additional conditions which lead to the identifiability of $(\frac{a'b}{b'}, b'm, c_0(\theta), c_1(\theta), c_2(\theta))$.

If $n = 3$, $\Lambda_3(\theta) = b'mF_2 + \frac{a'b}{b'}b'm(aF_1 + F_0) + c_3(\theta)$. Hence, F_0, F_1, F_2 being random variables, we get that $(a, \frac{a'b}{b'}, b'm)$ are identifiable. Now, identifying $(b\sigma, b'\tau, b'u)$ consists in solving a linear system using the conditions on $c_0(\theta), c_1(\theta), c_2(\theta))$. We obtain :

  - if $\{a \neq \frac{a'b}{b'}\}$, $(a, \frac{a'b}{b'}, b'm, b'u, b\sigma, b'\tau)$ is identifiable.

  - if $a = \frac{a'b}{b'}$, the identifiable parameters are $(a, b'm, b'u, b\sigma + b'\tau)$.

Noting that, for $i \geq 1$, $c_{i+1}(\theta) - ac_i(\theta) = (1 - a + \frac{a'b}{b'})b'u$ and that only $(a, b'm, \frac{a'b}{b'})$ enters in the F_i's coefficients. it can be checked that observing more generations does not lead to the idenfiability of additional parameters.

A.5. Proof of Proposition 4.2. Let $\phi = (a, \frac{a'b}{b'}, b'm, b'u, b\sigma, b'\tau)$ with $a \neq \frac{a'b}{b'}$. Under Asssumption (A7), we can define K_1, K_2, a_1, a_2 such that

$$\Phi \subset [K_1, K_2]^5 \times [a_1, a_2], \text{ with } 0 < K_1 < K_2 < +\infty \text{ and } 0 < a_1 < a_2 < 1.$$

The normalized loglikelihood process can be written as

(A.15) $$\frac{1}{K}\tilde{l}_K^1(\phi, Y_{0:n}(K)) = \frac{1}{K}\sum_{k=1}^{K} J(\phi, Y_{0:n}^k) \text{ with}$$

$$J(\phi, Y_{0:n}^k) = \sum_{i=0}^{n} R_i^k \log \Lambda_i(\phi, Y_{0:i-1}^k) - \Lambda_i(\phi, Y_{0:i-1}^k).$$

These random variables are i.i.d. and we have to study the behaviour of their empirical distribution with respect to ϕ. For getting the consistency of the associated maximum likelihood estimator, we have to prove that the parametric class $\{\mathbb{P}_\phi, \phi \in \Phi\}$ is Glivenko-Cantelli (see e.g. Van der Vaart, 1998). There is no close form for the density of these variables, so that no generic argument can here be applied: we thus propose a direct proof.

The functional $\{\phi \to J(\phi, Y_{0:n})\}$ is a.s twice continuously differentiable on Φ. Let $D\Lambda_i(\phi)$ denote the gradient in $\mathbb{R}^6$ of $\Lambda_i(\phi, Y_{0:i-1})$. Then,

$$\mid J(\phi'; Y_{0:n}) - J(\phi''; Y_{0:n}) \mid \leq \left(\sum_{i=0}^{n} R_i \sup_{\phi \in \Phi} \| (\frac{1}{\Lambda_i(\phi)} + 1)D\Lambda_i(\phi) \| \right) \| \phi' - \phi'' \| .$$

Under (A7), we have that, for all i, $\Lambda_i(\phi) \geq K_1 > 0$. Let us compute $D\Lambda_i(\phi)$ Using (4.4) for the definitions of $c_i(\theta)$ and $\Lambda_i(\theta)$, let us set

$$\Lambda_i(\phi) = b'mF_{i-1} + \frac{a'b}{b'}(b'm)(F_{i-2} + aF_{i-3} + \cdots + a^{i-2}F_0) + c_i(\phi).$$

For $i = 0$, we have

$$\frac{\partial\Lambda_0(\phi)}{\partial a} = \frac{\partial\Lambda_0(\phi)}{\partial(a'b/b')} = \frac{\partial\Lambda_0(\phi)}{\partial(b'm)} = \frac{\partial\Lambda_0(\phi)}{\partial(b'u)} = 0; \ \frac{\partial\Lambda_0(\phi)}{\partial(b\sigma)} = \frac{\partial\Lambda_0(\phi)}{\partial(b'\tau)} = 1.$$

Using the convention that non properly defined terms are set to 0 (e.g. the sum $(F_{i-3} + 2aF_{i-4} + \dots)$ is set to 0 for $i \leq 2$, we get, for $i \geq 1$,

$$\frac{\partial \Lambda_i(\phi)}{\partial a} = \frac{a'b}{b'}(b'm)(F_{i-3} + 2aF_{i-4} + \cdots + (i-2)a^{i-3}F_0) + \frac{\partial c_i(\phi)}{\partial a}, \text{ with}$$

$$\frac{\partial c_i(\phi)}{\partial a} = (b\sigma)ia^{i-1} + \frac{a'b}{b'}(b'u)\frac{d}{da}(\frac{1-a^{i-1}}{1-a}) + \frac{a'b}{b'}(b'\tau)(i-1)a^{i-2} ;$$

$$\frac{\partial \Lambda_i(\phi)}{\partial (a'b/b')} = b'm(F_{i-2} + aF_{i-3} + \cdots + a^{i-2}F_0) + (b'\tau)a^{i-1} + b'u\frac{1-a^{i-1}}{1-a} ;$$

$$\frac{\partial \Lambda_i(\phi)}{\partial (b'm)} = F_{i-1} + \frac{a'b}{b'}(F_{i-2} + aF_{i-1} + \cdots + a^{i-2}F_0) ;$$

$$\frac{\partial \Lambda_i(\phi)}{\partial (b'u)} = 1 + \frac{a'b}{b'}\frac{1-a^{i-1}}{1-a} ;$$

$$\frac{\partial \Lambda_i(\phi)}{\partial (b\sigma)} = a^i ;$$

$$\frac{\partial \Lambda_i(\phi)}{\partial (b'\tau)} = \frac{a'b}{b'}a^{i-1}.$$

All these partial derivatives are positive for $i \geq 1$ and, except the first one $\frac{\partial \Lambda_i(\phi)}{\partial a}$, they are bounded from above by $M_1(1 + \sum_{i=0}^{i-1} F_j)$ where M_1 is a constant determined by Φ (since on Φ, $0 < a_1 \leq a \leq a_2 < 1$). Noting now that the application $\{i \to ia^{i-1}\}$ satisfies $\forall i \geq 1$, $\alpha_1 < ia^{i-1} < \alpha_2$ on $[a_1, a_2] \subset (0,1)$ with $0 < \alpha_1 < \alpha_2 < +\infty$, we can also bound $\frac{\partial \Lambda_i(\phi)}{\partial a}$ by $M_2(1 + \sum_{j=0}^{i-3} F_j)$. Joining these bounds, we get,

$$\sup \| (\frac{1}{\Lambda_i(\phi)} + 1)D\Lambda_i(\phi) \| \leq M_3(1 + \sum_{j=0}^{i-1} F_j) ,$$

$$\mid J(\phi; Y_{0:n}) - J(\phi'; Y_{0:n}) \mid \leq \eta M_3 Z_n \text{ with } Z_n = \sum_{i=0}^{n}(1 + \sum_{j=0}^{i-1} F_j)R_i.$$

Using now that $\mathbb{E}_{\phi_0}(R_i \sum_{j=0}^{i-1} F_j) = \mathbb{E}_{\phi_0}(\Lambda_i(\phi_0) \sum_{j=0}^{i-1} F_j)$, Z_n satisfies,

$$\mathbb{E}_{\phi_0} Z_n \leq n\mathbb{E}_{\phi_0}\left(\sum_{i=0}^{n}(1 + F_j)^2\right) \leq n \sum_{i=0}^{n}\left(1 + \sqrt{\mathbb{E}_{\phi_0} F_i^2}\right)^2.$$

This is finite since $G(.)$ and $\mu(.)$ have finite variance and n is prescribed. Let us define the r.v. $Z_n^k = \sum_{i=0}^{n}(1 + \sum_{j=0}^{i-1} F_j^k)R_i^k$. Then the continuity modulus of $\tilde{l}_K^1(\theta)$ verifies :

$$w(K, \eta, \frac{1}{K}\tilde{l}_K^1) = \sup_{\|\phi'-\phi''\|\leq\eta} \frac{1}{K}\sum_{k=1}^{K} \mid J(\phi'; Y_{0:n}^k) - J(\phi''; Y_{0:n}^k) \mid \leq \eta\frac{1}{K}\sum_{k=1}^{K} Z_n^k.$$

Using now that $\mathbb{E}_{\phi_0} Z_n < \infty$, we can apply the strong law of large numbers to $w(K, \eta, \frac{1}{K} \tilde{l}^1_K)$, which is a sufficient condition for ensuring the consistency of $\hat{\phi}_K$.

It remains to study the asymptotic normality of the estimators. It is easy to check that the random variables $^\tau W^k = (W^k_1, \dots, W^k_6)$ with

$$W^k_p = \sum_{i=0}^{n} \left(\frac{R^k_i}{\Lambda^k_i(\phi_0)} - 1 \right) \frac{\partial \Lambda^k_i(\phi_0)}{\partial \phi_p}$$

are i.i.d. centered under $\mathbb{P}_{\phi_0}$, with covariance matrix Σ is for $1 \le p, q \le 6$

$$\Sigma_{p,q} = \Sigma_{p,q}(\phi_0) = \mathbb{E}_{\phi_0} \sum_{i=0}^{n} \frac{1}{\Lambda_i(\phi_0)} \frac{\partial \Lambda_i}{\partial \phi_p}(\phi_0) \frac{\partial \Lambda_i}{\partial \phi_q}(\phi_0).$$

This matrix is well-defined and finite is since, for all i, $\Lambda_i(\phi) \ge K_1 > 0$ and $\left| \frac{\partial \Lambda_i(\phi)}{\partial \phi_p} \right| \le M_3(1 + \sum_0^{i-1} F_j)$ for $p = 1, \dots 6$. Now, $\hat{\phi}_K$ is a zero for the score function, so that a Taylor expansion at ϕ_0 yields, using the consistency of the vector $\hat{\phi}_K$,

$$0 = \frac{1}{\sqrt{K}} D\tilde{l}^1_K(\hat{\phi}_K) = \frac{1}{\sqrt{K}} D\tilde{l}^1_K(\phi_0) + \frac{1}{K} \left(\nabla \tilde{l}^1_K(\phi_0) + R_K(\phi_0, \hat{\phi}_K) \right) \sqrt{K}(\hat{\phi}_K - \phi_0)$$

The term $\nabla \tilde{l}^1_K(\phi_0)$ contains the second derivatives of $\tilde{l}^1_K(\phi)$ w.r.t. ϕ_p, ϕ_q and $\tilde{R}_K(\phi_0, \hat{\phi}_K)$ is the remainder term of the Taylor expansion. So, the strong law of large numbers yields : $\frac{1}{K} \nabla \tilde{l}^1_K(\phi_0) \to - \sum_{i=0}^{n} \mathbb{E}_{\phi_0} \left(\frac{1}{\Lambda_i(\phi_0)} \frac{\partial \Lambda_i(\phi_0)}{\partial \phi_p} \frac{\partial \Lambda_i(\phi_0)}{\partial \phi_q} \right)$. The remainder term $R_K(\phi_0, \hat{\phi}_K)$ is bounded uniformly on Φ by $\| \hat{\phi}_K - \phi_0 \| Z_K$, with $Z_K = sup\{ \frac{1}{K} \nabla \tilde{l}^1_K(\phi), \phi \in \Phi \}$. Using that Z_K is bounded uniformly on Φ by $n^2 M_6 (1 + \sum_{i=0}^{n} F_i F_j)^2$, we get that $R_K(\phi_0, \hat{\phi})$ goes to 0 under $\mathbb{P}_{\phi_0}$, which leads to the result stated in Proposition 4.2 provided that $\Sigma(\phi_0)$ is invertible.

REFERENCES

Athreya, K. B. and Ney, P. E. (1972). *Branching Processes*. Springer-Verlag, Berlin.

Barry, S. C., Brooks, S. P., Catchpole, E. A., and Morgan, B. J. T. (2003). The analysis of ring-recovery data using random effects. *Biometrics* **59,** 54–65.

Bhat, B. R. and Adke, S. R. (1981). Maximum likelihood estimation for branching processes with immigration. *Advances in Applied Probability* **13,** 498–509.

Buckland, S. T., Newman, K. B., Thomas, L., and Koesters, N. B. (2004). State-space models for the dynamics of wild animal populations. *Ecological Modelling* **171,** 157–175.

Cappé, O., Moulines, E., and Ryden, T. (2005). *Inference in Hidden Markov Models.* Springer-Verlag, New York.

Caswell, H. (2001). *Matrix Population Models.* Sinauer Associates, Inc. Publishers, Sunderland, USA.

Claessen, D., Dilligan, C. A., Lutman, P. J., and van den Bosch, F. (2005a). Which traits promote persistence of feral gm crops? part 1: implications of environmental stochasticity. *Oikos* **110,** 20–29.

Claessen, D., Dilligan, C. A., Lutman, P. J., and van den Bosch, F. (2005b). Which traits promote persistence of feral gm crops? part 2: implications of metapopulation structure. *Oikos* **110,** 30–42.

Clark, J. and Bjornstad, O. (2004). Population time series: Process variability, observation errors, missing values, lags, and hidden states. *Ecology* **85,** 3140–3150.

Colbach, N., Clermont-Dauphin, C., and Meynard, J. M. (2001). Genesys : a model of the influence of cropping system on gene escape from herbicide tolerant rapeseed crops to rape volunteers- i. temporal evolution of a population of rapeseed volunteers in a field. *Agriculture, Ecosystems and Environment* **83,** 235–253.

Crawley, M. and Brown, S. (1995). Seed limitation and the dynamics of feral oilseed rape on the m25 motorway. *Proceedings of Royal Society of London B* **259,** 49–54.

Dacunha-Castelle, D. and Duflo, M. (1993). *Probabilités et Statistiques. 2.Problèmes à temps mobile.* Masson, 2eme edition.

David, O., Garnier, A., Larédo, C., and Lecomte, J. (2008). Estimation of plant demographic parameters from stage-censuses. in preparation.

de Valpine, P. (2004). Monte carlo state-space likelihoods by weighted posterior kernel density estimation. *Journal of the American Statistical Association* **99,** 523–535.

Dempster, A. P., Laird, N. M., and Rubin, D. B. (1977). Maximum likelihood fron incomplete data via the em algorithm. *Journal of the Royal Statistical Society Series B* **39,** 1–38.

Garnier, A. (2006). *Dynamique et dispersion d'une espèce cultivée échappée des champs: le cas du colza.* Ecologie, Université Paris-Sud 11.

Garnier, A., David, O., Lecomte, J., Deville, A., and Larédo, C. (2008). Are seed immigration and local seed production efficient to favour the persistence of feral oilseed rape populations submitted to weed control? in preparation.

Garnier, A., Deville, A., and Lecomte, J. (2006). Stochastic modelling of feral plant populations with seed immigration and road verge management. *Ecological Modelling* **197,** 373–382.

Genon-Catalot, V., Jeantheau, T., and Larédo, C. (2003). Conditional likelihood estimators for hidden markov models and stochastic volatility models. *Scandinavian Journal of Statistics* **30,** 297–316.

Genon-Catalot, V. and Larédo, C. (2006). Leroux's method for general hidden markov models. *Stochastic Processes and their Applications* **116,** 222–243.

Gonzalez, M., Martin, J., Martinez, R., and Mota, M. (2008). Non-parametric bayesian estimation for multitype branching processes through simulation-based methods. *Computational Statistics & data Analysis* **52,** 1281–1291.

Gross, K., Craig, B., and Hutchison, W. D. (2002). Bayesian estimation of a demographic matrix model from stage-frequency data. *Ecology* **83,** 3285–3298.

Guttorp, P. (1991). *Statistical Inference for Branching Processes.* Wiley, New York.

Haccou, P., Jagers, P., and Vatutin, V. A. (2005). *Branching processes: Variation, growth, and extinction of populations.* Cambridge University Press.

Hall, P. and Heyde, C. C. (1980). *Martingale Limit Theory and Its Applications.* Wiley, New York.

Kimmel, M. and Axelrod, D. E. (2002). *Branching Processes in Biology.* Springer-Verlag, New York.

Kuhn, E. and Lavielle, M. (2004). Coupling a stochastic approximation version of EM with an MCMC procedure. *ESAIM: P&S* **8,** 115–131.

Maaouia, F. and Touati, A. (2005). Identification of multitype branching processes. *The Annals of Statistics* **33,** 2655–2694.

McLachlan, G. J. and Krishnan, T. (2007). *The EM Algorithm and Extensions.* Wiley Series in Probability and Statistics. John Wiley & Sons, second edition.

Melbourne, B. A. and Hastings, A. (2008). Extinction risk depends stongly on factors contributiong to stochasticity. *Nature* **454,** 100–103.

Mode, C. (1971). *Multitype Branching Processes.* American Elsevier, New York.

Mode, C. J. and Sleeman, C. K. (2000). *Branching Processes in Epidemiology.* World Scientific, Singapore.

Olsson, J., Cappé, O., Douc, R., and Moulines, E. (2008). Sequential monte carlo smoothing to parameter estimation in non linear state space models. *Bernoulli* **64,** 155–179.

Pivard, S., Adamczyk, K., Lecomte, J., Lavigne, C., Bouvier, A., Deville, A., Gouyon, P. H., and Huet, S. (2008). Where do the feral oilseed rape populations come from? A large-scale study of their possible origin in a farmland area. *Journal of Applied Ecology* **45,** 476–485.

R Development Core Team (2005). *R: A language and environment for statistical computing.* R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0.

Raftery, A. E., Givens, G. H., and Zeh, J. E. (1995). Inference for a deterministic population dynamics model for bowhead whales. *Journal of the American Statistical Association* **90,** 402–430.

Royle, J. A. (2004). N-mixture models for estimating population size from spatially replicated counts. *Biometrics* **60,** 108–115.

Sung, Y. J. and Geyer, C. J. (2007). Monte carlo likelihood inference for missing data models. *The Annals of Statistics* **35,** 990–1011.

Thomas, L., Buckland, S. T., Newman, K. B., and Harwood, J. (2005). A unified framework for modelling wildlife population dynamics. *Australian and New Zealand Journal of Statistics* **47,** 19–34.

Van der Vaart, A. W. (1998). *Asymptotic Statistics.* Cambridge University Press.

Wei, C. Z. and Winnicki, J. (1990). Estimation of the means in the branching process with immigration. *The Annals of Statistics* **18,** 1757–1773.

Winnicki, J. (1991). Estimation of the variances in the branching process with immigration. *Probability Theory and Related Fields* **88,** 77–106.

Wood, S. N. (1997). *Inverse problems and structured populations dynamics*, pages 555–586. S. Turjapurkar and H. Caswell editors, Chapman & Hall, New York.

Catherine Larédo,
UR341, Mathématiques et Informatique appliquées,
INRA, F-78350 Jouy-en-Josas, France
& CNRS, UMR7599, Probabilités et Modèles aléatoires,
Univeristés Paris 6 & 7, 75005 Paris France.
E-mail: catherine.laredo@jouy.inra.fr

Olivier David,
UR341, Mathématiques et
Informatique appliquées,
INRA, F-78350 Jouy-en-Josas,France.
E-mail: olivier.david@jouy.inra.fr

Aurélie Garnier
Université Paris Sud,
Laboratoire Ecologie, Sytématique & Evolution,
UMR8079, F-91405 Orsay, France;CNRS, 91405 0rsay, France; AgroParisTech, 75231 Paris, France.
E-mail: au.garnier@gmail.com