

# Quantifying time-inhomogeneous stochastic introgression processes with hazard rates

Atiyo Ghosh<sup>a,\*</sup>, Maria Conceição Serra<sup>b</sup>, Patsy Haccou<sup>a,c</sup>

<sup>a</sup>*Institute of Environmental Sciences (CML), Leiden University, P.O. Box 9518, Leiden 2300 RA, The Netherlands*

<sup>b</sup>*Center of Mathematics, University of Minho, Campus de Gualtar, 4710-057 Braga, Portugal*

<sup>c</sup>*Mathematical Institute, Leiden University, P.O. Box 9512, Leiden 2300 RA, The Netherlands*

---

## Abstract

Introgression is the permanent incorporation of genes from one population into another through hybridization and backcrossing. It is currently of particular concern as a possible mechanism for the spread of modified crop genes to wild populations. The hazard rate is the probability per time unit that such an escape takes place, given that it has not happened before. It is a quantitative measure of introgression risk that takes the stochastic elements inherent in introgression processes into account. We present a methodology to calculate the hazard rate for situations with time-varying gene flow from a crop to a large recipient wild population. As an illustration, several types of time-inhomogeneity are examined, including deterministic periodicity as well as random variation. Furthermore, we examine the effects of an extended fitness bottleneck of hybrids and backcrosses in combination with time-varying gene flow. It is found that bottlenecks decrease the hazard rate, but also slow down and delay its changes in reaction to changes in gene flow. Furthermore, we find that random variation in gene flow generates a

lower hazard rate than analogous deterministic variation. We discuss the implications of our findings for crop management and introgression risk assessment.

*Keywords:* Branching Process, Invasion, Transgene, Risk management, Random environment

---

## 1 **1. Introduction**

2 Through backcrossing and hybridization, genes from one population can  
3 become permanently incorporated into the genome of another population.  
4 This process is called introgression (Riesberg and Wendel, 1993; Ellstrand  
5 et al., 1999; Hails and Morley, 2005). Introgression of crop genes into wild  
6 relatives may have severe negative environmental effects, such as the spread  
7 of insecticide or herbicide resistance genes. In particular, there are strong  
8 concerns about transgene escape and its consequences, e.g. the production  
9 of superweeds (Maan, 1987; Snow et al., 1999; Thompson et al., 2003; Kelly  
10 et al., 2005).

11 The likelihood of such scenarios, given environmental conditions, crop  
12 management, and characteristics of the species involved can be studied  
13 with mathematical models. Such models allow us to perform thought ex-  
14 periments, and identify factors that crucially determine introgression risk.  
15 Introgression usually involves many random components, such as hybridiza-  
16 tion and backcross events, and demographic stochasticity in hybrid popula-  
17 tions. In a previous paper (Ghosh and Haccou, 2010) we showed that it is  
18 important to take this stochasticity into account, since stochastic models

---

\*Corresponding author, Tel: +31715275608, Fax: +31715277434

*Email address:* [ghosh@cml.leidenuniv.nl](mailto:ghosh@cml.leidenuniv.nl) (Atiyo Ghosh)

*Preprint submitted to Theoretical Population Biology*

*November 28, 2012*

19 may give very different predictions from deterministic ones. We considered  
20 a situation where foreign genes invade repeatedly into a resident wild pop-  
21 ulation, and each invasion has a small probability of establishing a perma-  
22 nent lineage (see also Haygood et al., 2004). We showed that there can be  
23 an extensive period of failed invasions, and that the length of this period  
24 largely determines introgression risk. Furthermore, we derived a measure,  
25 the hazard rate, that quantifies the distribution of such periods. In the con-  
26 text of introgression, the hazard rate is defined as the probability per time  
27 unit that a permanent lineage is initiated, given that this has not happened  
28 before. It is derived from a multitype branching process model of hybrid  
29 population dynamics (Demon et al., 2007; Serra and Haccou, 2007).

30 In our previous paper we assumed that the distribution of numbers of  
31 newly created hybrids is the same in each time period. We considered  
32 a model with an initial fitness bottleneck (i.e. F1 hybrids have a lower  
33 fitness than the wild type) and showed that in such a situation, the hazard  
34 rate increases monotonically from zero to a constant asymptotic value. As  
35 a consequence, the distribution of the initial period before establishment  
36 of a permanent lineage can be approximated by a time-lagged geometric  
37 distribution. In many applications, however, the hybridization probability  
38 will vary in time, due to, for example, crop rotation or termination, or  
39 random variation, such as weather-dependent pollinator activity. In the  
40 current paper we generalize the method to include such time-inhomogeneity.  
41 We calculate the hazard rate for general time-inhomogeneous hybridization  
42 schemes and examine the effects of crop management schemes such as (gra-  
43 dually) stopping or increasing crop cultivation, or rotating crops. We show  
44 that, in the latter case, periods in which the hazard rate increases alternate

45 with periods of decrease, and that, in the long run, it converges to a periodic  
46 function. We also examine how stochastic fluctuations in hybridization rates  
47 affect the hazard rate.

48 As an example we consider a model for a monocarpic species (it dies  
49 after flowering), that is monoecious (flowers have both male and female  
50 functions), and non-selfing. We first consider a situation where F1 hybrids  
51 have a reduced fitness when compared to the wild-type, and all backcrosses  
52 have the same life history parameters, and superior fitness. Then the model  
53 is generalized to examine the effects of an extended fitness bottleneck, where  
54 several initial backcross generations have a reduced fitness.

55 There are many other contexts in which repeated invasions with low ini-  
56 tial fitness occur, such as tumor spread and growth, where usually several  
57 mutations must occur before cells proliferate (as in Michor et al., 2006), or  
58 pathogen host switching, where adjustments to new hosts imply an initial  
59 fitness bottleneck (as in Reluga et al., 2007). Time-inhomogeneity of in-  
60 vasions may play a role in such contexts too. For instance, there may be  
61 time-varying risks of exposure to carcinogenic environments (e.g. Bos et al.,  
62 2004). Furthermore, many epidemics show time-varying infection patterns  
63 (as in Welliver, 2009). Our methods and results therefore have implications  
64 for research in such contexts too.

## 65 **2. The model**

66 We consider a plant species that dies after flowering once. For simplicity,  
67 we assume that there is no age-dependence. Furthermore, it is assumed that  
68 there is a large, stable wild population, and random numbers of hybrid seeds  
69 are produced by pollen flow from a nearby crop. We consider time periods

70 of one year. Seeds may germinate at the beginning of the year, and plants  
71 grow up to be adults and may flower later in the same year. We denote the  
72 probability that a seed germinates and that the seedling survives to become  
73 an adult plant by  $p_0$ . In this paper we will consider  $p_0$  as a given parameter.  
74 Its value is determined by the population dynamics of the wild population,  
75 and is such that this population is stable (see Ghosh and Haccou, 2010, for  
76 an example of its calculation).

77 Hybrid formation can be followed by repeated backcrossing with wild  
78 plants. F1 hybrids are assumed to be less fit than wild individuals, but  
79 backcrossed individuals have a positive probability of producing a perma-  
80 nent introgressed lineage. We assume that all backcross generations are  
81 equivalent with respect to their life history parameters, and therefore they  
82 do not need to be distinguished as separate types (this assumption is re-  
83 laxed in section 6). As a consequence, there are two types of plants in the  
84 model: F1 hybrids (labelled type-1) and backcrossed individuals (labelled  
85 type- $E$ ).

86 Since the population of wild plants is large and the numbers of individ-  
87 uals containing crop genes are initially small, it can be assumed that these  
88 individuals do not interact with each other, but only with wild plants. This  
89 has several implications. Firstly, since we consider a non-selfing species,  
90 reproduction can only occur through outcrossing with wild plants. Sec-  
91 ondly, competition occurs only with the wild population. This is quantified  
92 through the probability  $p_0$ . For convenience, we assume that there are no  
93 other factors apart from this competition that affect germination proba-  
94 bility of hybrids and backcrosses. The model can be easily generalized to  
95 account for e.g. effects of spatial variation.

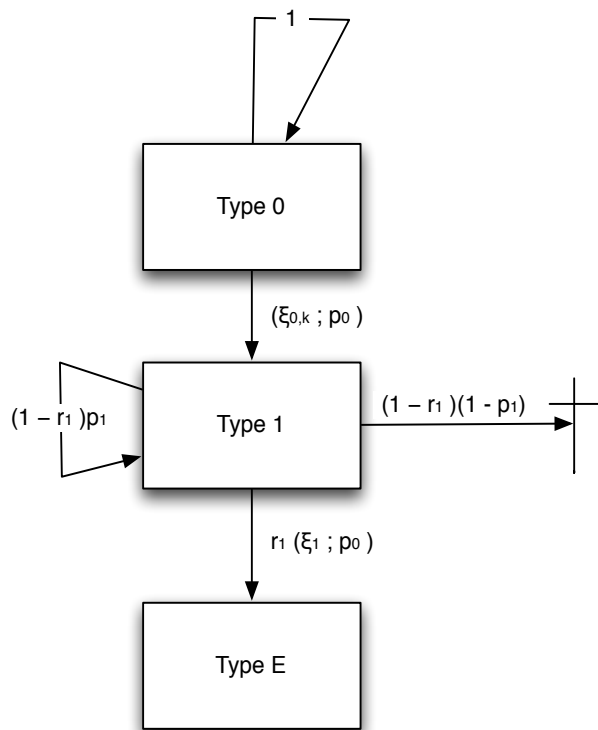


Figure 1: Schematic representation of the model.  $(\xi_{0,k}; p_0)$  and  $(\xi_1; p_0)$  represent the production of  $\xi_{0,k}$  and  $\xi_1$  seeds respectively, where each seed has a germination probability  $p_0$ . Each type- $E$  individual initiates a lineage which eventually becomes extinct with probability  $q$ .

96 Because hybrid and backcrossed plants do not affect each other's repro-  
 97 duction and survival initially, their invasion dynamics can be modeled as  
 98 a branching process. The production of hybrid seeds is modeled by means  
 99 of an artificial type, which we will call type-0. There is one permanently  
 100 present individual of this type, that produces a stochastic number of hy-  
 101 brid seeds in each year. Fig. 1 shows a schematic summary of the invasion  
 102 dynamics.

103 The model thus involves three different types of individuals: type-0, type-

104 1 and type- $E$ . Each year, a type-0 individual produces one individual of  
 105 type-0 and a random number of F1 hybrid seeds. In our previous paper  
 106 we assumed that the probability distribution of these random numbers was  
 107 the same over time. In this paper, we let it vary over years. The number  
 108 of hybrid seeds produced in year  $k$  is a random variable denoted by  $\xi_{0,k}$ .  
 109 Each one of these seeds germinates and produces a type-1 individual with  
 110 probability  $p_0$ . Type-1 individuals flower with probability  $r_1$ , and produce a  
 111 random number,  $\xi_1$ , of backcrossed seeds, either by male or female functions.  
 112 In the case that a type-1 individual does not flower (with a probability  
 113  $(1 - r_1)$ ), it may then survive to become a type-1 individual in the next  
 114 year with probability  $p_1$ , or it will die with a probability  $1 - p_1$ . Each  
 115 backcrossed seed germinates and survives with probability  $p_0$ , to produce  
 116 a type- $E$  individual. Type- $E$  individuals produce only type- $E$  offspring in  
 117 their lineage. We denote the probability that a lineage started by one type- $E$   
 118 individual goes extinct by  $q$ . This value can be calculated straightforwardly  
 119 from the life history parameters of type- $E$  individuals, by standard methods  
 120 (see e.g. Haccou et al., 2005; Ghosh and Haccou, 2010). Here, we will treat  
 121 it as a parameter in the model, taking values between zero and one.

### 122 3. Derivation of the hazard rate

123 Probability generating functions are important tools in deriving the haz-  
 124 ard rate. Let  $X$  be a non-negative discrete random variable, then its prob-  
 125 ability generating function (p.g.f.) is a function from  $[0, 1]$  to  $[0, 1]$  which  
 126 is defined as  $E[s^X]$ , where  $E[.]$  denotes expectation. The p.g.f. of  $\xi_{0,k}$  is  
 127 denoted by  $G_0(k; s)$ , and that of  $\xi_1$  by  $G_1(s)$ .

128 Define the random variable  $I_i(k, n)$  ( $n, k \in \mathbb{N}_0$ ,  $i = 0, 1$ ) to be the total

129 number of type- $E$  individuals with non type- $E$  parents, appearing up to  
 130 and including year  $n$ , in the *line of descent* of a single individual of type- $i$   
 131 that was produced in year  $k$ . The expression *line of descent* refers to the  
 132 population process stemming from the referred individual. For a general  
 133 scenario where individuals can have offspring of any type, this definition  
 134 leads to the following equalities:

$$I_i(k, n) = \begin{cases} 0 & \text{if } k \geq n \\ Z_E^{(i)}(k+1) + \sum_{m=0}^1 Z_m^{(i)}(k+1) \sum_{j=1} I_m^{(j)}(k+1, n) & \text{if } k < n \end{cases} \quad (1)$$

135 where  $Z_m^{(i)}(k+1)$  represents the number of type- $m$  individuals that the  
 136 type- $i$  individual (born in year  $k$ ) produced in year  $k+1$ . The  $I_m^{(j)}(k+1, n)$   
 137 terms represent the total number of type- $E$  individuals that have non type-  
 138  $E$  parents, appearing up to year  $n$  in the line of descent of the  $j^{\text{th}}$  individual  
 139 of type- $m$  that was born in year  $k+1$  from the initial type- $i$  individual.

In the specific scenario described in Fig. 1, we find the following recursive relationships in  $k$  for the different p.g.f.'s of the  $I_i(k, n)$ 's, where  $f_{I_i(k, n)}(s)$  denotes the p.g.f. of  $I_i(k, n)$  (see Appendix Appendix A.1):

$$\begin{aligned} f_{I_0(k, n)}(s) &= f_{I_0(k+1, n)}(s)G_0(k; p_0 f_{I_1(k+1, n)}(s) + 1 - p_0) \\ f_{I_1(k, n)}(s) &= (1 - r_1)(1 - p_1) + (1 - r_1)p_1 f_{I_1(k+1, n)}(s) + r_1 G_1(p_0 s + 1 - p_0) \end{aligned} \quad (2)$$

140 with the initial conditions  $f_{I_1(n, n)}(s) = f_{I_0(n, n)}(s) = 1$ . Note that, since the  
 141 seed production of type-1 individuals is homogeneous,

$$f_{I_1(k, n)}(s) = f_{I_1(0, n-k)}(s). \quad (3)$$

142 The time of an introgression event,  $T$ , is defined as the time that the first  
 143 type- $E$  individual is produced whose lineage never becomes extinct. The



144 population starts with a single type-0 individual, therefore:

$$P(T > n) = f_{I_0(0,n)}(q), \quad (4)$$

145 since the probability that an introgression event occurs after a time  $n$  is the  
 146 probability that all type- $E$  individuals produced at or before year  $n$  have  
 147 become extinct.

148 The hazard rate of introgression is defined as the probability per time  
 149 unit that an introgression event occurs given that it has not occurred before.  
 150 With time units of one year, this gives:

$$H_n(q) = P(T = n | T > n - 1) / \text{year} = \left( 1 - \frac{f_{I_0(0,n)}(q)}{f_{I_0(0,n-1)}(q)} \right) \text{year}^{-1} \quad (5)$$

151 with  $n \in \mathbb{N}_0$ .

152 The second equation of (2) can be solved to yield (see Appendix Appendix  
 153 A.2):

$$f_{I_1(0,n)}(s) = 1 - \beta_1(s) + \beta_1(s) b_1^n, \quad (6)$$

154 where, in order to simplify future expressions, we have introduced the quan-  
 155 tities

$$b_1 = (1 - r_1) p_1 \quad \text{and} \quad \beta_1(s) = \frac{r_1 (1 - G_1(p_0 s + 1 - p_0))}{1 - b_1} \quad (7)$$

156 Putting (2), (3), (4) and (5) together gives us the following expression for  
 157 the hazard rate (see Appendix Appendix A.3):

$$H_n(q) = \begin{cases} 0 & \text{if } n \in \{0, 1\} \\ 1 - \frac{\prod_{j=1}^{n-1} G_0(j-1; p_0 f_{I_1(0,n-j)}(q) + 1 - p_0)}{\prod_{j=1}^{n-2} G_0(j-1; p_0 f_{I_1(0,n-1-j)}(q) + 1 - p_0)} & \text{if } n \geq 2 \end{cases} \quad (8)$$

158 which can be computed by using (6). This result provides us with a general  
 159 method for calculating the hazard rate with time-inhomogeneous hybridiza-  
 160 tion. In the next sections we examine several situations.

#### 161 4. Deterministically varying hybridization

162 For mathematical convenience we assume that hybrids are generated ac-  
 163 cording to a Poisson distribution with a time-dependent mean, i.e.:

$$G_0(k; s) = e^{-m_0(k)(1-s)}, \quad s \in [0, 1]. \quad (9)$$

164 We also take  $\xi_1$  as Poisson-distributed with mean  $m_1$  in presented numerical  
 165 work.

166 Combining (6) to (9) gives:

$$H_n(q) = \begin{cases} 0 & \text{if } n \in \{0, 1\} \\ 1 - e^{-p_0 \beta_1(q)(1-b_1)b_1^{n-2} \sum_{j=0}^{n-2} m_0(j) b_1^{-j}} & \text{if } n \geq 2. \end{cases} \quad (10)$$

From (10) it follows that the long term behaviour of the hazard rate depends on the limit behaviour, as  $k \rightarrow \infty$ , of:

$$b_1^k \sum_{j=0}^k \frac{m_0(j)}{b_1^j}.$$

167 For example, if  $m_0(j) = m_0^j$ , the hazard rate converges to zero when  
 168  $0 < m_0 < 1$  and it converges to one when  $m_0 > 1$ . If there is constant  
 169 hybridization, i.e  $m_0(j) = m_0$ , the hazard rate tends to a constant value  
 170 between zero and one (as was also derived in Ghosh and Haccou, 2010). It  
 171 can easily be shown that, for the current model, this value equals

$$1 - \exp\{-p_0 \beta_1(q) m_0\}. \quad (11)$$

172 In the next subsections we will examine the effects of specific frequently  
 173 used crop-management schemes.

#### 174 4.1. Temporary crops

175 Crop cultivation may be stopped for a variety of reasons. In the case of  
 176 transgene crops, e.g., legislation may change, or termination of cultivation  
 177 may be used as a management strategy to lower the chance of introgression.  
 178 In this sub-section we examine the case where hybridization occurs at a  
 179 constant rate, and is then stopped at a fixed time  $S$ , i.e.:

$$m_0(j) = \begin{cases} m_0 & \text{if } 0 \leq j < S \\ 0 & \text{if } j \geq S, \end{cases} \quad (12)$$

180 with  $m_0 > 0$ .

181 Substituting this into (10) gives:

$$H_n(q) = \begin{cases} 0 & \text{if } n \in \{0, 1\} \\ 1 - e^{-m_0 p_0 \beta_1(q) (1-b_1^{n-1})} & \text{if } 2 \leq n \leq S + 1 \\ 1 - e^{-m_0 p_0 \beta_1(q) b_1^{n-(S+1)} (1-b_1^S)} & \text{if } n \geq S + 2 \end{cases} \quad (13)$$

182 Thus, the hazard rate increases monotonically to a maximum level of  
 183  $1 - e^{-m_0 p_0 \beta_1(q) (1-b_1^S)}$  at time  $S + 1$  and decays monotonically afterwards.  
 184 The decay is only seen to start at time  $S + 2$  because stopping hybridiza-  
 185 tion at year  $S$  will only affect the population of type-1 individuals at time  
 186  $S + 1$ , and the population of type- $E$  individuals at time  $S + 2$ . The rate  
 187 of increase as well as that of decay is mainly governed by  $b_1$ , which rep-  
 188 represents the probability that individuals do not flower but do survive (see  
 189 (7)). A larger value of  $b_1$  makes the hazard rate increase and decrease more  
 190 slowly. When  $b_1$  tends to zero (i.e. when the probability of flowering in the

191 first year is high and/or the survival probability of non-flowering adults is  
 192 low), the maximum level is reached quickly and, unless  $S$  is very small, it  
 193 is therefore virtually independent of  $S$ . Furthermore, after stopping culti-  
 194 vation, the hazard rate returns rapidly to zero. As  $b_1$  tends to zero or  $S$   
 195 tends to infinity, the maximum level approaches the asymptotic level of the  
 196 hazard rate in the situation without stopping. The effect of the life history  
 197 parameters on this asymptotic level can be inferred from (11).

198 With temporary crops, there is a positive probability that introgression  
 199 never occurs. From (4), (9), (12) and the derivation in Appendix Appendix  
 200 A.3 it is apparent that this probability equals:

$$\lim_{n \rightarrow \infty} P(T > n) = \lim_{n \rightarrow \infty} f_{I_0(0,n)}(q) = e^{-m_0 p_0 \beta_1(q) S} \quad (14)$$

201 Thus, it decreases exponentially with the stopping time  $S$ , at a rate deter-  
 202 mined by the hybridization rate and the life history parameters.

203 A numerical example of the shape of the hazard rate for two different  
 204 stopping times (10 and 20 years) is given in Fig. 2a. In this example,  
 205 the hazard rate increases quickly, and, as a consequence, its maximum level  
 206 does not noticeably differ for the two chosen stopping times. The probability  
 207 distribution of  $T$  can be expressed in terms of the hazard rate as follows  
 208 (see e.g. Kalbfleisch and Prentice, 2002):

$$P(T = x) = \prod_{i=0}^{x-1} (1 - H_i(q)) H_x(q). \quad (15)$$

209 For small values of  $H_n(q)$ , the product term is close to one, and the proba-  
 210 bility becomes nearly equal to the hazard rate. This is demonstrated in Fig.  
 211 2b. As can be seen from the figure, the probabilities of introgression events  
 212 happening quite early are relatively large, i.e. the probability distributions

213 are very skewed, similar to the situation with constant crop cultivation ex-  
 214 amined before in Ghosh and Haccou (2010). For the numerical examples in  
 215 Fig. 2b, the probabilities that no introgression occurs at all are respectively  
 216 0.985 ( $S = 10$ ) and 0.970 ( $S = 20$ ).

#### 217 4.2. Crop rotation

218 Crop rotation is often used to maintain soil quality and prevent the build  
 219 up of pathogens. It may also be used as a management strategy to lower  
 220 introgression risk. In this section we study the situation where periods  
 221 with hybridization at a constant rate alternate with periods without hy-  
 222 bridization. The duration of hybridization periods is denoted by  $S$ , and the  
 223 durations of the *hybridization pauses* by  $R$ . Thus we have:

$$m_0(j) = \begin{cases} m_0 & \text{if } v(R + S) \leq j < v(R + S) + S \\ 0 & \text{if } v(R + S) + S \leq j < (v + 1)(R + S) \end{cases} \quad (16)$$

224 with  $v \in \mathbb{N}_0$ .

225 It can be shown (see Appendix Appendix A.4) that in the long run the  
 226 hazard rate tends to a periodic function with period  $R + S$ , i.e. if we define  
 227 the time:

$$k = n - v(R + S) - 2 \quad (17)$$

228 then, for  $n$  tends to infinity the hazard rate becomes:

$$\mathcal{H}_k(q) = \begin{cases} 1 - e^{-m_0 p_0 \beta_1(q) \left(1 - b_1^{k+1} \frac{1 - b_1^R}{1 - b_1^{R+S}}\right)} & \text{if } 0 \leq k < S \\ 1 - e^{-m_0 p_0 \beta_1(q) b_1^{k+1-S} \frac{(1 - b_1^S)}{1 - b_1^{(R+S)}}} & \text{if } S \leq k < R + S \end{cases} \quad (18)$$

229 The time in (17) is the time after the  $v$ th crop rotation shifted by two  
 230 time units. The shift of two units is for mathematical convenience, and  
 231 corresponds for the first two years where the hazard rate is zero.

232 This result implies that periods in which instantaneous introgression risk  
 233 is high alternate with periods in which it is low. Figure 2c illustrates that  
 234 this asymptotic behavior can be reached very quickly. Figure 2d shows the  
 235 corresponding probabilities of introgression events happening at time  $x$ . As  
 236 noted previously, the probability distribution is nearly equal to the hazard  
 237 rate initially, but (inevitably) decreases with  $x$ .

238 There are different ways to quantify the effect of a given crop rotation  
 239 scheme on the hazard rate. The asymptotic maximum hazard rate can be  
 240 found by substituting  $k = S - 1$  in (18), leading to:

$$1 - e^{-m_0 p_0 \beta_1(q) \frac{1-b_1^S}{1-b_1^{(R+S)}}}, \quad (19)$$

241 and the minimum by substituting  $k = R + S - 1$ , which gives:

$$1 - e^{-m_0 p_0 \beta_1(q) b_1^R \frac{1-b_1^S}{1-b_1^{(R+S)}}}. \quad (20)$$

242 For the numerical example in Figure 2c the asymptotic maximum hazard  
 243 rate equals 0.00154, and the minimum is of the order  $10^{-6}$ . As can be seen  
 244 from the figure, these values are reached quite soon.

245 An alternative measure is the long-run average hazard rate. This is found  
 246 by fitting the survivor function of a constant hazard rate to the survivor  
 247 function of the hazard rate from (18). This approach leads to the following  
 248 value for the long-run average hazard rate (see Appendix Appendix A.5 for  
 249 details):

$$\lambda \approx 1 - e^{-p_0 m_0 \beta_1(q) \frac{S}{R+S}}. \quad (21)$$

250 Thus, the long-run average hazard rate is the same as the asymptotic hazard  
 251 rate with a continuous crop and a constant expected number of newly pro-  
 252 duced hybrids equal to  $S/(R + S)$  times  $m_0$ . In Fig. 2d we have indicated

253 the time-distributions corresponding to a continuous immigration with the  
254 maximum hazard rate (c.f. (19)) and the long-run average hazard rate.

## 255 5. Randomly varying hybridization

256 Until now we have considered deterministic variation in hybridization  
257 rates. In many cases, however, there will also be random variation. For  
258 instance, weather conditions will vary over different years, and this may  
259 affect pollen dispersal from the crop to local wild populations. Such random  
260 variations can be independent, or (positively or negatively) autocorrelated.  
261 In this section, we consider the effect of random variation according to  
262 different regimes.

263 Random temporal variation of  $m_0$  can be included in the model by using  
264 different type-0 individuals. Thus, we consider  $\gamma$  different types, denoted  
265 by type-(0,  $i$ ) ( $i = 1, \dots, \gamma$ ). A type-(0,  $i$ ) individual produces a number of  
266 type-1 seeds according to a p.g.f.  $G_{0,i}(s)$ , and with probability  $\kappa_{i,j}$  also  
267 exactly one individual of type-(0,  $j$ ) ( $j = 1, \dots, \gamma$ ), so  $\sum_{j=1}^{\gamma} \kappa_{i,j} = 1$  for all  $i$ .

268 As an illustration, consider the case where the environment alternates  
269 between two states according to a two-type Markov chain. In that case  
270  $\gamma = 2$ . When the environment is state 1, a Poisson-distributed number of  
271 hybrids is formed, i.e.  $G_{0,1}(s) = e^{-m_0(1-s)}$  and when the environment is in  
272 state 2, no hybrids are produced, i.e.  $G_{0,2}(s) = 1$ . The transition probability  
273 from state 1 to state 2 equals  $\kappa_{1,2}$  and that from state 2 to state 1 equals  
274  $\kappa_{2,1}$ . An independently varying environment corresponds to the situation  
275 where  $\kappa_{1,2} + \kappa_{2,1} = 1$ . In the case of positive autocorrelation, this sum is  
276 smaller than one whereas it is larger than one for negatively autocorrelated  
277 environments.

278 As a special case, consider an independently varying environment, with  
 279  $\kappa_{1,1} = \kappa_{2,1} = S/(R+S)$  and  $\kappa_{1,2} = \kappa_{2,2} = R/(R+S)$ . Note that the expected  
 280 proportion of years with positive hybridization numbers is the same as in  
 281 the crop rotation scenario considered in (16). We assume that the process  
 282 is stationary. The hazard rate is then given by (see Appendix Appendix  
 283 A.6)

$$H_n(q) = \frac{S}{R+S} \left( 1 - e^{-m_0 p_0 (1 - f_{I_1(0,n-1)}(q))} \right). \quad (22)$$

284 Using the solution of  $f_{I_1(0,n)}(q)$  from (6) and taking large  $n$  leads to the  
 285 asymptotic value:

$$H_\infty(q) = \frac{S}{R+S} [1 - e^{-m_0 p_0 \beta_1(q)}]. \quad (23)$$

286 To examine the effects of autocorrelation, let  $\kappa_{1,2} = \kappa_{2,1} = 1 - \kappa_{1,1} =$   
 287  $1 - \kappa_{2,2} = \alpha$ . The environment is negatively autocorrelated if  $\alpha > 0.5$ ,  
 288 positively autocorrelated if  $\alpha < 0.5$ , and independent if  $\alpha = 0.5$ . The equa-  
 289 tions given in Appendix Appendix A.6 can be used to calculate the hazard  
 290 rate for these models numerically . Figure 3a shows the resulting asymp-  
 291 totic hazard rate for different values of  $\alpha$ . As can be seen, there is not  
 292 much difference between negatively autocorrelated or independent environ-  
 293 ments. The asymptotic hazard rate is much reduced, however, when there  
 294 is a strong positive autocorrelation. With this choice of parameters, the  
 295 probability of a year with hybridization is  $1/2$ , and so the situation is com-  
 296 parable to a crop rotation scenario with  $S = R$ , as in Fig. 2(c). Note that  
 297 the situation where  $\alpha = 1$  corresponds to deterministic alternation between  
 298 one-year periods with and without a positive hybridization probability. In  
 299 this scenario, the hazard rate still approaches an asymptotic hazard rate  
 300 because the process is initiated by the stationary-distribution of type-(0, 1)



301 and type-(0, 2) individuals, as depicted in Fig. 3b. In a specific realisation,  
302 the hazard rate then oscillates as previously observed, which is also shown  
303 in Fig. 3b, where the process is initiated by a single type-(0, 1) individual.

## 304 6. Effects of bottlenecks

305 Until now we have considered the situation where all backcrossed genera-  
306 tions are more fit than the wild type. However, often there is outbreeding  
307 depression, which implies that several backcrosses are needed before a fitness  
308 advantage is observed (e.g. Edmands, 2002). In this section we extend the  
309 model to account for such situations, and investigate effects of the length  
310 of the bottleneck on the hazard rate.

311 The generalized model involves  $L + 2$  ( $L \in \mathbb{N}$ ) different types: types  
312  $0, 1, \dots, L$ , and type- $E$ . Type-0 individuals are defined as before. The flow-  
313 ering probability of type- $i$  ( $i \in \{1, 2, \dots, L\}$ ) is denoted by  $r_i$ , the p.g.f. of  
314 their seed production by  $G_i(s)$  and their seeds will produce type- $(i + 1)$   
315 adults. The survival probability of non-flowering type- $i$  individuals is  $p_i$ ,  
316 and survivors remain of type  $i$ . The offspring of type- $L$  individuals will be  
317 of type- $E$ . Type- $E$  individuals and  $q$  are defined as in previous sections.  
318 The scheme is represented in Fig. 4.

319 The hazard rate in this scenario follows a similar method to the derivation  
320 in the previous case, but see Appendix Appendix A.7 for full details. Nu-  
321 merical solutions of the supremum of the hazard rate against  $L$  are shown  
322 in Fig. 5a for the crop-rotation situation described in (16).

323 To further examine the effect of bottlenecks, we consider a Taylor ap-  
324 proximation of the hazard rate around the point  $q = 1$ , for the case that  
325 plants are annual (i.e.  $r_i = 1$  for  $i = 1, 2, \dots, L$ ). The resulting Taylor

326 approximation is (see Appendix A.8 for details):

$$H_n(q) \approx \left( p_0 m_0 (n - L - 1) \prod_{i=1}^L p_0 m_i \right) (1 - q) \quad (24)$$

327 where  $m_i$ ,  $i = 1, 2, \dots, L$ , represents the average number of seeds produced  
328 by a type- $i$  individual.

329 When the values of  $m_i$  are similar, this expression decreases geometrically  
330 with  $L$ , which corresponds to the shape observed in Fig. 5a.

331 Bottlenecks not only reduce the maximum hazard rate, but also induce  
332 a delay in the changes of the hazard rate in reaction to changes in crop  
333 cultivation. This is illustrated in Fig. 5b.

## 334 7. Discussion

335 In this paper we generalize our previous results on hazard rates of in-  
336 trogression (Ghosh and Haccou, 2010) to situations with time-varying hy-  
337 bridization. Whereas in our previous paper we considered a model with two  
338 age classes and a bottleneck of one generation, the present paper concerns  
339 situations without age dependence, and effects of extended bottlenecks. The  
340 general methodology that we present can be extended straightforwardly to  
341 other types of life histories. Furthermore, there are several general conclu-  
342 sions that are valid for a wide range of situations.

343 First of all, the results shed light on the meaning of the hazard rate as a  
344 measure of stochastic introgression rate, and its practical implications. As  
345 illustrated in this paper, hazard rates may increase and decrease in time, in  
346 relation to changes in the magnitude of hybridization rates. When the hy-  
347 bridization rate is high, the instantaneous risk of introgression events is also  
348 high. During such periods, increased vigilance is advisable, to prevent the

349 successful establishment of crop genes in wild populations. When hybridiza-  
350 tion frequency drops, the hazard rate decreases, and accordingly, vigilance  
351 might be decreased. Our results show, however, that managers must take  
352 care not to let their guards down too soon, since increased fitness bottle-  
353 necks delay the changes in the hazard rate. This implies, for instance, that  
354 even after crop cultivation has been terminated for a considerable time, the  
355 risk of introgression events may still be quite high (see Fig. 5b), reaffirming  
356 a conclusion from Haygood et al. (2003).

357 The risk that introgression occurs is determined by the interaction be-  
358 tween life history and fitness characteristics of hybrids, and crop manage-  
359 ment. As we illustrated, changes in gene flow induce changes in the level  
360 of the hazard rate. The speed at which such changes take place, as well  
361 as the magnitude of the hazard rate depends on life-history characteristics.  
362 For instance, increases in fitness bottlenecks not only cause a delay in ad-  
363 justment of the hazard rate, but also decelerate the adjustments, and lower  
364 the maximum level. Furthermore, in all scenarios, the maximum level of  
365 the hazard rate is affected by the factor  $\beta_1(q)$ , which is determined by the  
366 fitness of the backcrosses (see (7)).

367 We examined the effect of several possible scenarios. With temporary  
368 crops, there is a positive probability that introgression does not occur, that  
369 depends on the duration of the crop cultivation. Furthermore, in this situ-  
370 ation, the hazard rate at a given time  $x$  is nearly equal to the probability  
371 of an introgression event at that time, and thus provides a good approxi-  
372 mation for the probability distribution (see e.g. Fig. 2b). This is a general  
373 result, that can be derived from the relation between the hazard rate and  
374 the time-distribution.

375 With crop rotation, the hazard rate becomes periodic, and fluctuations  
376 also occur in the time-distribution of introgression events (Fig. 2c and d).  
377 In such situations, a simpler measure of risk might sometimes be needed.  
378 One option is to use the hazard rate that in the long run would lead to  
379 the same introgression risk over a given period as the crop rotation scheme.  
380 This value is given in (21), and indicated in Fig. 2c. We refer to this  
381 value as the long-run average hazard rate. However, please note that it  
382 is not the same as the arithmetic time-average of the asymptotic hazard  
383 rate. From (21) it can be seen that the average risk level is determined  
384 by the proportion of years that crop cultivation occurs. Thus, the average  
385 hazard rate remains the same when  $S$  and  $R$  are multiplied by the same  
386 factor. For instance, alternating between one year 'on' and 'off' would in  
387 the long run give the same average hazard rate as alternating between, say,  
388 ten years 'on' and 'off'. Larger values of  $S$  and  $R$  would, however, lead to  
389 a larger amplitude of the fluctuations in the hazard rate. The magnitude  
390 of this effect can be calculated by means of (19) and (20). In situations  
391 with large fluctuations the use of the average hazard rate as a risk indicator  
392 might be misleading, since the maximum hazard rate is much higher than  
393 the average. This is illustrated in Fig. 2c. In such a situation, the time-  
394 distribution of introgression events corresponding to the average hazard rate  
395 is also radically different from the real one (see Fig. 2d).

396 Another possible way to quantify the risk is to use the long-run maximum  
397 hazard rate, which provides a conservative measure of risk. Figure 2d also  
398 shows the time-distribution of introgression events corresponding to the  
399 maximum hazard rate, illustrating that in an example with large amplitude  
400 of the hazard rate this might be a better risk measure.

401 We also derived methods to calculate the hazard rate in situations with  
402 randomly varying hybridization rates. As a specific example, we considered  
403 a situation where the environment alternates between two states, one with  
404 and one without hybridization, according to a Markov chain. In the absence  
405 of environmental autocorrelation, the hazard rate becomes constant in the  
406 long run, and an explicit expression can be derived. This value is given  
407 in (23), and corresponds to the arithmetic time-average of the asymptotic  
408 hazard rate in a deterministic crop rotation scheme with the same propor-  
409 tion of years of hybridization as the random environment. It can be shown  
410 that this value is lower than the long-run average hazard rate given in (21).  
411 Therefore, random variation in gene flow appears to reduce the probability  
412 that introgression occurs. This also appears to be true in autocorrelated  
413 environments, as illustrated in Fig.3. Positive autocorrelation reduces the  
414 hazard rate, whereas negative autocorrelation does not seem to have much  
415 effect. In any case, the long-run hazard rate is smaller than the long-run  
416 average for the deterministically alternating environment. Thus, we expect  
417 that hazard rates for deterministic scenarios provide conservative measures  
418 for introgression risk. This is a fortunate result, since in many situations  
419 there is likely to be random variation in gene flow, which is beyond control  
420 of management measures.

421 We examined several specific gene flow scenarios, to illustrate the method-  
422 ology and its possibilities. For mathematical tractability, we used a rela-  
423 tively simple life-history and Poisson distributions for the numbers of hy-  
424 brids. Our methods can readily be adjusted to examine other types of gene  
425 flow variation, more complicated life histories, and hybrid number distri-  
426 butions. In such cases, however, no explicit expressions for (asymptotic)

427 hazard rates will be possible. Instead, numerical methods will have to be  
428 used, based on the adjusted equations. Such calculations generally do not  
429 take much time on a standard computer.

430 Other generalizations, which are the subject of ongoing research, include  
431 the introduction of time-inhomogeneity in backcross fitness, multi-locus ge-  
432 netics, and meta-population dynamics. Another type of generalization con-  
433 cerns small populations. As long as wild receptor populations are assumed  
434 to be large enough to exclude direct interactions between initial invaders,  
435 the approach that we used up to now, based on branching processes, can be  
436 applied. For small populations however, different methods need to be de-  
437 veloped, based on density-dependent models (see (e.g Jagers and Klebaner,  
438 2000)). This is another line of ongoing research.

439 The use of stochastic models in introgression studies is quite rare, al-  
440 though not completely absent (e.g. Haygood et al., 2004; Thompson et al.,  
441 2003). The general methodology for handling such models, and quantifying  
442 introgression timing events is, however, still in its infancy. The use of hazard  
443 rates is, in our opinion, an important step forward. Serra and Haccou (2007)  
444 introduced the concept of the hazard rate for studying branching processes  
445 with mutation, and Ghosh and Haccou (2010) were the first to use it in the  
446 context of introgression. The work presented here represents the next step  
447 of a research program that is aimed at developing a full-fledged toolbox for  
448 studying stochastic introgression processes. Such tools are indispensable  
449 in introgression risk management, since stochastic elements are inevitably  
450 present, and, furthermore, adding stochasticity changes the features of in-  
451 trogression processes considerably.

452 **8. Acknowledgments**

453 This research was funded through the research program 'Ecology Re-  
 454 garding Genetically modified Organisms (ERGO)', commissioned by four  
 455 Dutch ministries. This funding program is managed by the Earth and Life  
 456 Sciences Council (ALW) of the Netherlands Organisation for Scientific Re-  
 457 search (NWO). P. Haccou's research is additionally supported by the NDNS  
 458 (Nonlinear Dynamics of Natural Systems) program of NWO. M.C. Serra  
 459 would like to thank the Fundação para a Ciência e Tecnologia for finan-  
 460 cial support through the scholarship SFRH/BPD/47615/2008. We thank  
 461 Marije Stoops and Prof. Baorong Lu for discussions and comments on a  
 462 previous version.

463 **Appendix A. Appendix**

464 *Appendix A.1. Derivation of (2)*

Using (1) and the definition of p.g.f.'s we find:

$$\begin{aligned}
 f_{I_i(k,n)}(s) &= E \left[ E \left[ s^{I_i(k,n)} | Z_0^{(i)}(k+1), Z_1^{(i)}(k+1), Z_E^{(i)}(k+1) \right] \right] \\
 &= E \left[ E[s^{I_0(k+1,n)}]^{Z_0^{(i)}(k+1)} E[s^{I_1(k+1,n)}]^{Z_1^{(i)}(k+1)} E[s]^{Z_E^{(i)}(k+1)} \right] \\
 &= E \left[ f_{I_0(k+1,n)}(s)^{Z_0^{(i)}(k+1)} f_{I_1(k+1,n)}(s)^{Z_1^{(i)}(k+1)} s^{Z_E^{(i)}(k+1)} \right] \quad (\text{A.1})
 \end{aligned}$$

465 We can manipulate (A.1) as above because the individual lineages are in-  
 466 dependent of each other, and individuals of the same type have identical  
 467 offspring distributions.

468 Now we introduce the joint p.g.f of the reproduction distribution of a  
 469 type- $i$  individual belonging to a year  $k$  which, for  $i \in \{0, 1\}$  and  $k \geq 0$ , is

470 defined as

$$F_i(k; (s_0, s_1, s_E)) = E \left[ s_0^{Z_0^{(i)}(k+1)} s_1^{Z_1^{(i)}(k+1)} s_E^{Z_E^{(i)}(k+1)} \right] \quad (\text{A.2})$$

471 for  $(s_0, s_1, s_E) \in [0, 1]^3$ .

472 Putting (A.1) and (A.2) together, we find that

$$f_{I_i(k,n)}(s) = F_i(k; (f_{I_0(k+1,n)}(s), f_{I_1(k+1,n)}(s), s)) \quad (\text{A.3})$$

473 In our specific model, we have the following reproduction laws:

$$F_0(k; (s_0, s_1, s_E)) = s_0 G_0(k; p_0 s_1 + (1 - p_0)) \quad (\text{A.4})$$

474

$$F_1(k; (s_0, s_1, s_E)) = (1 - r_1)(1 - p_1) + (1 - r_1)p_1 s_1 + r_1 G_1(p_0 s_E + 1 - p_0). \quad (\text{A.5})$$

475 Substituting (A.5) and (A.4) into (A.3) gives (2).

476 *Appendix A.2. Derivation of (6)*

477 Since the population initiated by a type-1 individual is time-  
 478 homogeneous,  $f_{I_1(k,n)}(s) = f_{I_1(0,n-k)}(s)$ . Using this in the second equation  
 479 of (2) results in:

$$f_{I_1(0,n-k)}(s) = (1 - r_1)(1 - p_1) + (1 - r_1)p_1 f_{I_1(0,n-k-1)}(s) + r_1 G_1(p_0 s + 1 - p_0) \quad (\text{A.6})$$

Introducing  $b_1 = (1 - r_1)p_1$  and  $a_1(s) = (1 - r_1)(1 - p_1) + r_1 G_1(p_0 s + 1 - p_0)$ ,  
 allowing  $k = 0$ , this can be rewritten as follows:

$$\begin{aligned} f_{I_1(0,n)}(s) &= a_1(s) + b_1 f_{I_1(0,n-1)}(s) \\ &= a_1(s) + b_1 (a_1(s) + b_1 f_{I_1(0,n-2)}(s)) \\ &= \dots \\ &= b_1^n + a_1(s) \sum_{i=0}^{n-1} b_1^i. \end{aligned} \quad (\text{A.7})$$



480 Computing the geometric sum above, and taking the quantities defined in  
 481 (6) gives the required result.

482 *Appendix A.3. Derivation of (8)*

Deriving (8) follows from repeating equation (2) in the following way:

$$\begin{aligned}
 f_{I_0(0,n)}(s) &= f_{I_0(1,n)}(s)G_0(0; p_0 f_{I_1(1,n)}(s) + 1 - p_0) \\
 &= f_{I_0(2,n)}(s)G_0(1; p_0 f_{I_1(2,n)} + 1 - p_0)G_0(0; p_0 f_{I_1(1,n)}(s) + 1 - p_0) \\
 &\vdots \\
 &= \prod_{j=1}^{n-1} G_0(j-1; p_0 f_{I_1(0,n-j)}(s) + 1 - p_0)
 \end{aligned} \tag{A.8}$$

483 The expression in (8) follows from substituting (A.8) into (5).

484 *Appendix A.4. Derivation of (18)*

485 Substituting (16) into (10) gives the hazard rate. During the  $(v+1)^{th}$   
 486 period that hybridization is introduced, i.e. if  $v(R+S) + 2 \leq n < v(R+S) + S + 2$ ,  
 487 the following holds:

$$H_n(q) = 1 - e^{-m_0 p_0 \beta_1(q) \left( 1 - b_1^{n-(1+v(S+R))} + b_1^{n-(S+1)} (1 - b_1^S) \left( \frac{1 - b_1^{v(R+S)}}{b_1^{(v-1)(R+S)} (1 - b_1^{R+S})} \right) \right)} \tag{A.9}$$

488 and for the  $(v+1)^{th}$  period that hybridization is stopped, i.e. if  $v(R+S) +$   
 489  $S + 2 \leq n < (v+1)(R+S) + 2$ ,

$$H_n(q) = 1 - e^{-m_0 p_0 \beta_1(q) b_1^{n-(S+1)} (1 - b_1^S) \left( \frac{1 - b_1^{(v+1)(R+S)}}{b_1^{v(R+S)} (1 - b_1^{R+S})} \right)} \tag{A.10}$$

490 and, as in (10), the hazard rate equals zero for  $n \in \{0, 1\}$ . Substituting (17)  
 491 into (A.9) leads to the following for  $0 \leq k < S$  :

$$H_{v(R+S)+2+k}(q) = 1 - e^{-m_0 p_0 \beta_1(q) \left( 1 - b_1^{k+1} + b_1^{v(R+S)+k+1-S} (1 - b_1^S) \left( \frac{1 - b_1^{v(R+S)}}{b_1^{(v-1)(R+S)} (1 - b_1^{(R+S)})} \right) \right)} \quad (\text{A.11})$$

492 and substituting (17) into (A.10) leads to, for  $S \leq k < S + R$ :

$$H_{v(R+S)+2+k}(q) = 1 - e^{-m_0 p_0 \beta_1(q) b_1^{v(R+S)+k+1-S} (1 - b_1^S) \left( \frac{1 - b_1^{(v+1)(R+S)}}{b_1^{v(R+S)} (1 - b_1^{(R+S)})} \right)}. \quad (\text{A.12})$$

493 To reach the asymptotic behaviour described in (18), take  $v \rightarrow \infty$  in both  
 494 (A.11) and (A.12).

495 *Appendix A.5. Derivation of (21)*

496 First, note that the survival function of  $T$  and the hazard rate are related  
 497 as follows. For any  $t \in [0, +\infty)$ :

$$P[T > t] = \prod_{j \in \mathbb{N}_0 : j \leq t} (1 - H_j(q)). \quad (\text{A.13})$$

498 Define the sequence  $\{c_n, n \in \mathbb{N}_0\}$ :

$$c_n = \frac{P[T > n + R + S]}{P[T > n]}. \quad (\text{A.14})$$

499 The use of (A.8) with (4) and (A.14), gives:

$$\begin{aligned}
c_n &= \frac{f_{I_0(0,n+R+S)}(q)}{f_{I_0(0,n)}(q)} \\
&= \frac{\prod_{i=1}^{n+R+S-1} G_0(i-1; p_0 f_{I_1(0,n+R+S-i)}(q) + 1 - p_0)}{\prod_{i=1}^{n-1} G_0(i-1; p_0 f_{I_1(0,n-i)}(q) + 1 - p_0)} \\
&= \frac{\prod_{i=1}^{R+S} e^{-p_0 m_0(i-1)(1-f_{I_1(0,n+R+S-i)}(q))} \prod_{i=R+S+1}^{n+R+S-1} e^{-p_0 m_0(i-1)(1-f_{I_1(0,n+R+S-i)}(q))}}{\prod_{i=1}^{n-1} e^{-p_0 m_0(i-1)(1-f_{I_1(0,n-i)}(q))}} \\
&= e^{-p_0 m_0 \sum_{i=1}^S (1-f_{I_1(0,n+R+S-i)}(q))}
\end{aligned} \tag{A.15}$$

500 Note how the second product in the numerator is identical to the denomi-  
501 nator. This is a result of the periodicity of the hybridization rate in (16).  
502 Also, note that for  $S+1 \leq i \leq R+S$ ,  $m(i) = 0$ , which is used to reduce  
503 the number of terms in the sum.

504 When  $n \rightarrow \infty$ ,  $c_n$  converges to

$$C = e^{-p_0 m_0 S \beta_1(q)}. \tag{A.16}$$

505 Thus, in the long run, a process with a constant hazard rate,  $\lambda$ , and such  
506 that  $\lim_{n \rightarrow \infty} \frac{P[T > n+R+S]}{P[T > n]} = C$ , would have the same probability of an intro-  
507 gression event occurring within a period from  $n$  to  $n+R+S$ , with sufficiently  
508 large  $n$ . Using (A.13) and (A.16) we find that  $\lambda$  must satisfy

$$\lim_{n \rightarrow \infty} \prod_{i=n+1}^{n+R+S} (1 - \lambda) = C, \tag{A.17}$$

509 and the required result follows by combining (A.16) and (A.17) and solving  
510 for  $\lambda$ .

511 *Appendix A.6. Derivation of (22)*

512 Take the definitions of  $f_{I_i(k,n)}(s)$ ,  $I_i(k,n)$  and  $Z_m^{(i)}$  as before, but extend  
 513 it to include  $i = (0,1)$  and  $(0,2)$ . As before, a joint p.g.f. of the offspring  
 514 distribution of a single type- $i$  ( $i = (0,1), (0,2), 1, E$ ) is defined:

$$F_i(k; (s_{0,1}, s_{0,2}, s_1, s_E)) = E \left[ s_{0,1}^{Z_{0,1}^{(i)}(k+1)} s_{0,2}^{Z_{0,2}^{(i)}(k+1)} s_1^{Z_1^{(i)}(k+1)} s_E^{Z_E^{(i)}(k+1)} \right] \quad (\text{A.18})$$

515 Then, following the same methodology established in Appendix A.1, we get:

$$f_{I_i(k,n)}(s) = F_i(k; (f_{I_{0,1}(k+1,n)}(s), f_{I_{0,2}(k+1,n)}(s), f_{I_1(k+1,n)}(s), s)) \quad (\text{A.19})$$

Following further the methodology in Appendix A.1, the following recursive relationships hold:

$$f_{I_{0,1}(0,n-k)}(s) = G_{0,1} (p_0 f_{I_1(0,n-k-1)}(s) + 1 - p_0) \times (\kappa_{1,1} f_{I_{0,1}(0,n-k-1)}(s) + \kappa_{1,2} f_{I_{0,2}(0,n-k-1)}(s)) \quad (\text{A.20})$$

$$f_{I_{0,2}(0,n-k)}(s) = G_{0,2} (p_0 f_{I_1(0,n-k-1)}(s) + 1 - p_0) \times (\kappa_{2,1} f_{I_{0,1}(0,n-k-1)}(s) + \kappa_{2,2} f_{I_{0,2}(0,n-k-1)}(s)) \quad (\text{A.21})$$

where the simplifying expression  $f_{I_i(k,n)}(s) = f_{I_i(0,n-k)}(s)$  has been applied. Using the forms of  $G_{0,1}(s)$  and  $G_{0,2}(s)$  as specified in section 5, and setting  $k = 0$ , gives:

$$f_{I_{0,1}(0,n)}(s) = e^{-m_0 p_0 (1 - f_{I_1(0,n-1)}(s))} \times (\kappa_{1,1} f_{I_{0,1}(0,n-1)}(s) + \kappa_{1,2} f_{I_{0,2}(0,n-1)}(s))$$

$$f_{I_{0,2}(0,n)}(s) = \kappa_{2,1} f_{I_{0,1}(0,n-1)}(s) + \kappa_{2,2} f_{I_{0,2}(0,n-1)}(s) \quad (\text{A.22})$$

516 Since the environmental process is stationary:

$$P(T > n) = \frac{\kappa_{2,1}}{\kappa_{1,2} + \kappa_{2,1}} f_{I_{0,1}(0,n)}(q) + \frac{\kappa_{1,2}}{\kappa_{1,2} + \kappa_{2,1}} f_{I_{0,2}(0,n)}(q), \quad (\text{A.23})$$

517 and the hazard rate can be calculated from this. For the considered  
518 analog of the deterministic process without autocorrelation,  $f_{I_{0,1}(0,n)}(s) =$   
519  $e^{-m_0 p_0 (1 - f_{I_1(0,n-1)}(s))} f_{I_{0,2}(0,n)}(s)$ . Using (5) (A.23) and (A.22) then gives the  
520 required result.

521 *Appendix A.7. Derivation of the hazard rate in the bottleneck scenario*

522 We start by defining the random variable  $I_i(k, n)$  as before, except with  
523  $i \in \{0, 1, \dots, L\}$ . Also, we define p.g.f.'s,  $f_{I_i(k,n)}(s)$ , of these random vari-  
524 ables in the same way as previously done.

525 Since an individual belonging to a generation greater than  $n$  can produce  
526 no type- $E$  individuals before  $n$ , write the following for any  $i \in \{0, 1, \dots, L\}$ ,

$$I_i(k, n) = 0, \quad \text{if } k \geq n. \quad (\text{A.24})$$

527 Let us now turn to the case  $k < n$ . For a fixed  $i \in \{0, \dots, L\}$ , and  
528 a general scenario, where individuals can have offspring of any type, the  
529 following decomposition holds

$$I_i(k, n) = Z_E^{(i)}(k+1) + \sum_{m=0}^L Z_m^{(i)}(k+1) \sum_{j=1}^L I_m^{(j)}(k+1, n), \quad (\text{A.25})$$

where the random variables

$$Z_0^{(i)}(k+1), Z_1^{(i)}(k+1), \dots, Z_L^{(i)}(k+1), Z_E^{(i)}(k+1)$$

represent the number of offspring of types  $0, 1, \dots, L, E$ , respectively, that  
the initial type  $i$  produced. Also, as the notation suggests, the random  
variables

$$I_0^{(j)}(k+1, n), \quad j = 1, \dots, Z_0^{(i)}(k+1),$$

represent the number of type- $E$  individuals with non-type- $E$  parents, appearing up to and including year  $n$ , in the line of descent of the  $j^{\text{th}}$  type-0 offspring of the initial type- $i$  individual. Notice that, since the initial type- $i$  individual belongs to year  $k$ , its offspring belongs to year  $k+1$ . The random variables

$$\begin{aligned} I_1^{(j)}(k+1, n), & \quad j = 1, \dots, Z_1^{(i)}(k+1), \\ I_2^{(j)}(k+1, n), & \quad j = 1, \dots, Z_2^{(i)}(k+1), \\ & \quad \vdots \\ I_L^{(j)}(k+1, n) & \quad j = 1, \dots, Z_L^{(i)}(k+1), \end{aligned}$$

530 are defined in an analogous way, but now for the type-1, type-2, ..., type- $L$ ,  
531 respectively, offspring of the initial type- $i$  individual.

First manipulate the generating functions of (A.25) as follows:

$$\begin{aligned} f_{I_i(k,n)}(s) &= E \left[ E \left[ s^{I_i(k,n)} \mid Z_0^{(i)}(k+1), Z_1^{(i)}(k+1), \dots, Z_L^{(i)}(k+1), Z_E^{(i)}(k+1) \right] \right] \\ &= E \left[ E[s^{I_0(k+1,n)}]^{Z_0^{(i)}(k+1)} E[s^{I_1(k+1,n)}]^{Z_1^{(i)}(k+1)} \dots E[s^{I_L(k+1,n)}]^{Z_L^{(i)}(k+1)} E[s]^{Z_E^{(i)}(k+1)} \right] \\ &= E \left[ f_{I_0(k+1,n)}(s)^{Z_0^{(i)}(k+1)} f_{I_1(k+1,n)}(s)^{Z_1^{(i)}(k+1)} \dots f_{I_L(k+1,n)}(s)^{Z_L^{(i)}(k+1)} s^{Z_E^{(i)}(k+1)} \right] \end{aligned} \tag{A.26}$$

532 We can manipulate (A.26) as above because the individual lineages are  
533 independent of each other, and individuals of the same type have identical  
534 offspring distributions.

535 Introduce the joint p.g.f of the reproduction distribution of a type- $i$  in-  
536 dividual belonging to a year  $k$  which, for  $i \in \{0, 1, \dots, L\}$  and  $k \geq 0$ , is  
537 defined as

$$F_i(k; (s_0, s_1, \dots, s_L, s_E)) = E[s_0^{Z_0^{(i)}(k+1)} s_1^{Z_1^{(i)}(k+1)} \dots s_L^{Z_L^{(i)}(k+1)} s_E^{Z_E^{(i)}(k+1)}] \tag{A.27}$$

538 for  $(s_0, s_1, \dots, s_L, s_E) \in [0, 1]^{L+2}$ .

539 Putting (A.26) and (A.27) together, we find that

$$f_{I_i(k,n)}(s) = F_i(k; (f_{I_0(k+1,n)}(s), f_{I_1(k+1,n)}(s), \dots, f_{I_L(k+1,n)}(s), s)) \quad (\text{A.28})$$

540 In our specific model, we have the following assumptions regarding the  
541 reproduction:

- 542 • the reproduction law of a type 0 individual depends on the year number  
543 and the corresponding p.g.f. is given by

$$F_0(k; (s_0, s_1, \dots, s_L, s_E)) = s_0 G_0(k; p_0 s_1 + (1 - p_0)) \quad (\text{A.29})$$

- 544 • for a type  $i$  individual, with  $i \in \{1, \dots, L\}$ , the reproduction law does  
545 not depend on the year number and the corresponding p.g.f. is given  
546 by

$$\begin{aligned} F_i(k; (s_0, s_1, \dots, s_i, s_{i+1}, \dots, s_L, s_E)) &= (1 - r_i)(1 - p_i) + (1 - r_i)p_i s_i \\ &+ r_i G_i(p_0 s_{i+1} + 1 - p_0) \end{aligned} \quad (\text{A.30})$$

with  $s_{L+1} \equiv s_E$ . The fact that the reproduction law of these individuals is independent of time implies that

$$f_{I_i(k,n)}(s) = f_{I_i(0,n-k)}(s).$$

547 This relation will be used more or less explicitly in the following calcu-  
548 lations.

The use of (A.30) and (A.28) with  $i = L$ , gives

$$f_{I_L(0,n)}(s) = (1 - r_L)(1 - p_L) + (1 - r_L)p_L f_{I_L(0,n-1)}(s) + r_L G_L(p_0 s + 1 - p_0).$$

549 The use of initial condition  $f_{I_L(0,0)}(s) = 1$  results in the following for any  
 550  $n \geq 0$ , which is :

$$f_{I_L(0,n)}(s) = 1 - \beta_L(s) + \beta_L(s) b_L^n, \quad (\text{A.31})$$

551 with

$$b_L = (1 - r_L) p_L \quad \text{and} \quad \beta_L(s) = \frac{r_L (1 - G_L(p_0 s + 1 - p_0))}{1 - b_L}. \quad (\text{A.32})$$

552 The calculation of (A.31) above follows the same reasoning shown in Ap-  
 553 pendix Appendix A.2.

554 Now that we can calculate the p.g.f.'s of  $I_L(0, n)$ , we proceed by finding  
 555 expressions for the p.g.f.'s of  $I_i(0, n)$  for  $i = 0, 1, \dots, L - 1$ .

Note that, in the line of descent of a single type- $i$  individual belonging to year 0, new type- $E$  individuals can only appear after  $L - i + 1$  years (this is intuitively clear from Fig. 4). Hence, for  $i \in \{1, \dots, L - 1\}$ ,

$$f_{I_i(0,1)}(s) = f_{I_i(0,2)}(s) = \dots = f_{I_i(0,L-i)}(s) = 1.$$

Now, for  $n > L - i$ , the use of (A.30) and (A.28), gives

$$f_{I_i(0,n)}(s) = (1 - r_i)(1 - p_i) + r_i G_i(p_0 f_{I_{i+1}(0,n-1)}(s) + 1 - p_0) + (1 - r_i) p_i f_{I_i(0,n-1)}(s).$$

556 Repeating the procedure gives

$$\begin{aligned} f_{I_i(0,n)}(s) &= [(1 - r_i) p_i]^{n-(L-i)} + (1 - p_i) \sum_{j=1}^{n-(L-i)} (1 - r_i)^j p_i^{j-1} \\ &\quad + \sum_{j=1}^{n-(L-i)} r_i [(1 - r_i) p_i]^{j-1} G_i(p_0 f_{I_{i+1}(0,n-j)}(s) + 1 - p_0). \end{aligned}$$

557 Computing the sums above gives us the following p.g.f.'s:

$$f_{I_i(0,n)}(s) = 1 - \alpha_i + \alpha_i b_i^{n-(L-i)} + r_i \sum_{k=L-i}^{n-1} b_i^{n-k-1} G_i(p_0 f_{I_{i+1}(0,k)}(s) + 1 - p_0), \quad (\text{A.33})$$



558 where

$$b_i = (1 - r_i)p_i \quad \text{and} \quad \alpha_i = \frac{r_i}{1 - b_i}. \quad (\text{A.34})$$

559 We have  $f_{I_0(0,n)}(s) = 1$  for  $n \leq L$ , since a type-0 individual requires at  
 560 least  $L$  generations to produce a type-E individual. For  $n > L$  we combine  
 561 (A.29) and (A.28) to give:

$$f_{I_0(0,n)}(s) = \prod_{j=1}^{n-L} G_0(j-1; p_0 f_{I_1(0,n-j)}(s) + 1 - p_0) \quad (\text{A.35})$$

562 which can be calculated using (A.33) and (A.31).

563 The use of (A.35) and noting that, as before,  $P(T > n) = f_{I_0(0,n)}(q)$   
 564 yields the hazard rate:

$$H_n(q) = \begin{cases} 0 & \text{if } 0 \leq n \leq L \\ 1 - \frac{\prod_{j=1}^{n-L} G_0(j-1; p_0 f_{I_1(0,n-j)}(q) + 1 - p_0)}{\prod_{j=1}^{n-1-L} G_0(j-1; p_0 f_{I_1(0,n-1-j)}(q) + 1 - p_0)} & \text{if } n \geq L + 1. \end{cases} \quad (\text{A.36})$$

565 *Appendix A.8. Derivation of (24)*

Taking  $r_1 = 1$  in (A.31) to (A.34) gives:

$$f_{I_L(0,n)}(s) = 1 - \beta_L(s) \quad (\text{A.37})$$

$$f_{I_i(0,n)}(s) = G_i(p_0 f_{I_{i+1}(0,n-1)}(s) + 1 - p_0) \quad (\text{A.38})$$

where  $i = 1, 2, \dots, L - 1$ . Differentiating these expressions with respect to  $s$  and evaluating the results at the point  $s = 1$  gives:

$$\begin{aligned} f'_{I_L(0,n)}(1) &= p_0 m_L \\ f'_{I_i(0,n)}(1) &= p_0 m_i f'_{I_{i+1}(0,n-1)}(1) \end{aligned} \quad (\text{A.39})$$

566 where we have used the fact that the derivative of a p.g.f. evaluated at one  
 567 is the mean of the random variable.

Taking logarithms in (A.35) and differentiating at  $s = 1$  yields the following expression:

$$\begin{aligned}
 f'_{I_0(0,n)}(1) &= \sum_{j=1}^{n-L} p_0 m_0 (j-1) f'_{I_1(0,n-j)}(1) \\
 &= \sum_{j=1}^{n-L} p_0 m_0 (j-1) p_0^L \prod_{i=1}^L m_i
 \end{aligned}
 \tag{A.40}$$

568 where the last equality uses the expressions in (A.39).

569 Consider the representation of the hazard rate in (5). It is apparent that  
 570 the constant-term in the Taylor approximation will be zero, due to the fact  
 571 that p.g.f.'s evaluated at one are one. Taking the derivative of (5) around  
 572 one yields:

$$H'_n(1) = f'_{I_0(0,n-1)}(1) - f'_{I_0(0,n)}(1).
 \tag{A.41}$$

573 Using the above with (A.40) gives the required result.

574 Bos, P.M.J., Jan, P-J, van Raaij, M.T.M., 2004. Risk assessment of peak exposure to  
 575 genotoxic carcinogens: a pragmatic approach. *Toxicol. Lett.* 151, 43-50.

576 Demon, I., Haccou, P., van den Bosch, F., 2007. Introgression of resistance genes between  
 577 population: A model study of insecticide resistance in *Bemisia tabaci*. *Theor. Popul.*  
 578 *Biol.* 72, 292-304.

579 Edmands, S., 2002. Does parental divergence predict reproductive compatibility? *Trends*  
 580 *Ecol. Evol.* 17, 520-527.

581 Ellstrand, N.C., Prentice, H.C., Hancock, J.F., 1999. Gene flow and introgression from  
 582 domesticated plants into their wild relatives. *Annu. Rev. Ecol. Systemat.* 30, 539-563.

583 Ghosh, A., Haccou, P., 2010. Quantifying stochastic introgression processes with hazard  
 584 rates. *Theor. Popul. Biol.* 77, 171-180.

585 Haccou, P., Jagers, P., Vatutin, V.A., 2005. *Branching Processes: Variation Growth and*  
 586 *Extinction of Populations*. Cambridge University Press, Cambridge.

- 587 Hails, R.S., Morley, K., 2005. Genes invading new populations: A risk assessment per-  
588 spective. *Trends Ecol. Evol.* 20, 245-252.
- 589 Haygood, R., Ives, A.R., Andow, D.A., 2003. Consequences of recurrent gene flow from  
590 crops to wild relatives. *Proc. Biol. Sci.* 270, 1879-1896.
- 591 Haygood, R., Ives, A.R., Andow, D.A., 2004. Population genetics of transgene contain-  
592 ment. *Ecol. Lett.* 7, 213-220.
- 593 Jagers, P., Klebaner, F.C., 2000. Population-size-dependent and age-dependent branch-  
594 ing processes. *Stoch. Proc. Appl.* 87, 235-254.
- 595 Kalbfleisch, J.D., Prentice, R.L., 2002. *The Statistical Analysis of Failure Time Data*,  
596 2nd ed., John Wiley & Sons, New York.
- 597 Kelly, C.K., Bowler, M.J., Breden, F., Fenner, M., Poppy, G.M., 2005. An analytical  
598 model assessing the potential threat to natural habitats from insect resistance trans-  
599 genes. *Proc. Biol. Sci.* 272, 1759-1767.
- 600 Maan, S.S., 1987. Interspecific and intergeneric hybridisation in wheat. In: Heyne, E.G.  
601 (Ed.), *Wheat and Wheat Improvement*. ASA, CSSA and SSSA, Madison, pp. 453-461.
- 602 Michor, F., Nowak, M.A., Iwasa, Y., 2006. Stochastic dynamics of metastasis formation.  
603 *J. Theoret. Biol.* 240, 521-530.
- 604 Reluga, T., Meza, R., Walton, D.B., Galvani, A.P., 2007. Reservoir interactions and  
605 disease emergence. *Theor. Popul. Biol.* 72, 400-408.
- 606 Riesberg, L.H., Wendel, J.F., 1993. Introgression and its consequences in plants. In:  
607 Harrison, R.G. (Ed.), *Hybrid Zones and the Evolutionary Process*. Oxford University  
608 Press, Oxford, pp. 70-109.
- 609 Serra, M.C., Haccou, P., 2007. Dynamics of escape mutants. *Theor. Popul. Biol.* 72,  
610 167-178.
- 611 Snow, A.A., Andersen, B., Jorgensen, R.B., 1999. Costs of transgenic herbicide resistance  
612 introgressed from *Brassica napus* into weed *B. Rapa*. *Mol. Ecol.* 8, 605-615.
- 613 Thompson, C.J., Thompson, B.J.P., Ades, P.K., Cousens, R., Carinier-Gere, P., Land-  
614 man, K., Newbigin, E., Burgman, M.A., 2003. Model-based analysis of the likelihood  
615 of gene introgression from genetically modified crops into wild relatives. *Ecol. Model.*  
616 162, 199-209.
- 617 Welliver, R., 2009. The relationship of meteorological conditions to the epidemic activity

618 of respiratory syncytial virus. *Paediatr. Respir. Rev.*, 10 (2009) 6-8.

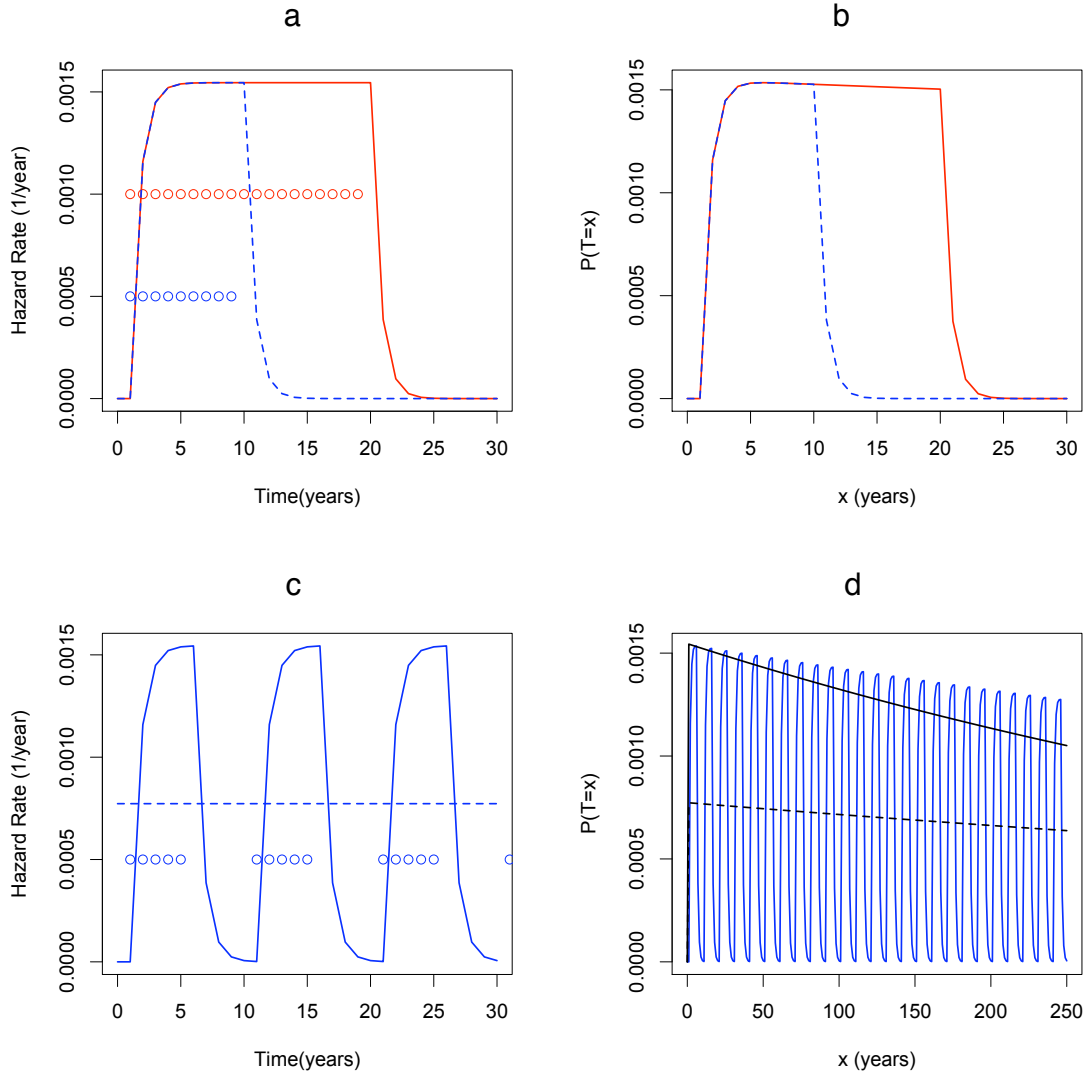


Figure 2: (a) Hazard rates when crops are terminated after a period of  $S = 10$  (blue), or  $S = 20$  (red). Parameter values:  $m_0 = 50$ ,  $p_0 = 0.001$ ,  $p_1 = r_1 = 0.5$ ,  $m_1 = 950$ ,  $q = 0.95$ , (b) The distributions of time until an introgression event, corresponding to the situations in (a). (c) The hazard rate with crop rotation (see (A.9) and (A.10)) (solid line) for  $R = S = 5$  and all other relevant parameters the same as in (a). The average hazard rate (see (21)) (dotted line). (d) Distribution of times until an introgression event for the crop rotation scenario of (c) (blue line), for a constant average hazard rate (dotted black line), and for a constant maximum hazard rate (see (19), solid black line). In (a) and (c), circles indicate periods when hybridization occurs, but not the amount of immigration.

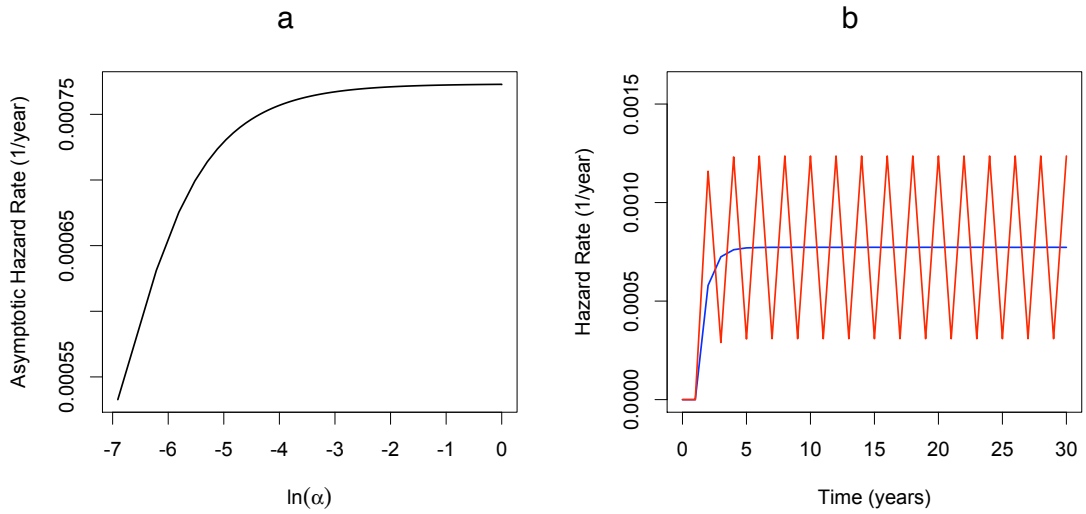


Figure 3: (a) The effect of autocorrelation on the asymptotic hazard rate when  $k_{1,2} = k_{2,1} = \alpha = 1 - k_{2,2} = 1 - k_{1,1}$ ,  $m_{0,1} = 50$ ,  $m_{0,2} = 0$ , and other parameter values as in Fig.2. The environment is positively autocorrelated when  $\ln \alpha < \ln 2 (\approx -0.69)$  and negatively autocorrelated when  $\ln \alpha > \ln 2$ . Periods with and without positive hybridization probabilities alternate deterministically when  $\ln \alpha = 0$ . (b) The hazard rate at  $\alpha = 1$  when the process is started with a stationary distribution of type-(0, 1) and type-(0, 2) individuals (blue), and when the process is started with a single type-(0, 1) individual (red).

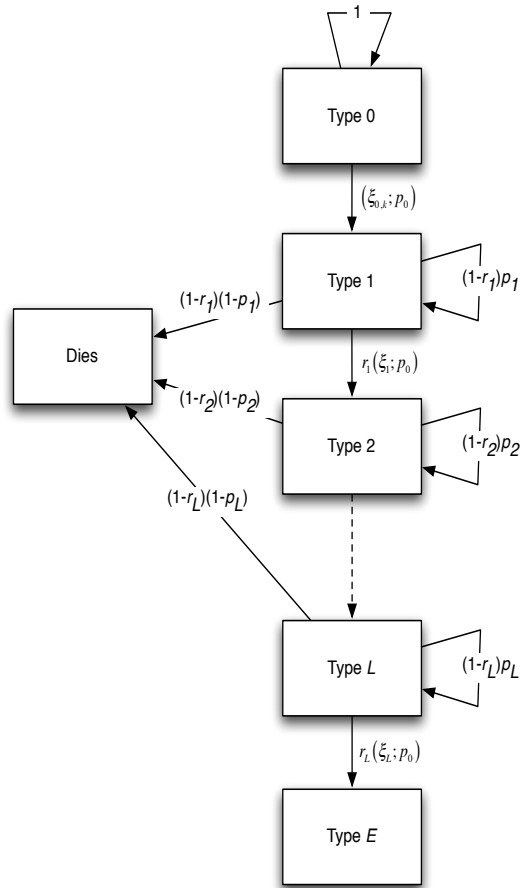


Figure 4: Schematic representation of the bottleneck model.  $(\xi_i; p_0)$  represents the production of  $\xi_i$  seeds  $i \in (0, k) \cup \{1, 2, \dots, L\}$ , where each seed has a germination probability  $p_0$ . Each type- $E$  individual initiates a lineage which eventually becomes extinct with probability  $q$ .

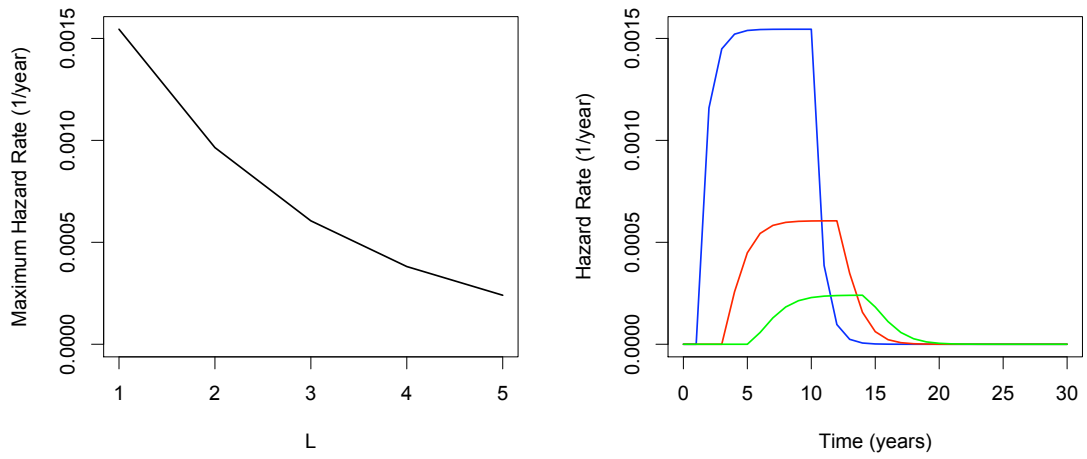


Figure 5: (a) The maximum hazard rate as a function of the length of the bottleneck  $L$  for a crop rotation scenario with  $R = S = 5$ ,  $m_0 = 50$ ,  $p_0 = 0.001$ ,  $p_i = r_i = 0.5$ ,  $m_i = 950$  for  $i = 1, 2, \dots, L$  and  $q = 0.95$ . (b) The hazard rate against time with hybridization as described in (12) with  $S = 10$  and all other parameters as in (a). The behaviour for  $L = 1$  (blue)  $L = 3$  (red) and  $L = 5$  (green) is shown.