

# *A polyetic modelling framework for plant disease emergence*

Article

Accepted Version

Willoquet, L., Savary, S., McDonald, B. A. and Mikaberidze, A. (2020) A polyetic modelling framework for plant disease emergence. *Plant Pathology*, 69 (9). pp. 1630-1643. ISSN 0032-0862 doi: 10.1111/ppa.13249 Available at <https://centaur.reading.ac.uk/91045/>

It is advisable to refer to the publisher's version if you intend to cite from the work. See [Guidance on citing](#).

To link to this article DOI: <http://dx.doi.org/10.1111/ppa.13249>

Publisher: Wiley-Blackwell

All outputs in CentAUR are protected by Intellectual Property Rights law, including copyright law. Copyright and IPR is retained by the creators or other copyright holders. Terms and conditions for use of this material are defined in the [End User Agreement](#).

[www.reading.ac.uk/centaur](http://www.reading.ac.uk/centaur)

**CentAUR**

Central Archive at the University of Reading

Reading's research outputs online

# 1 A polyetic modelling framework for plant disease emergence

2

3 Laetitia Willocquet<sup>a1,\*</sup>, S. Savary<sup>1</sup>, B.A. McDonald<sup>2</sup>, A. Mikaberidze<sup>2,3</sup>

4

5 <sup>1</sup>AGIR, INRAE, Université de Toulouse, INPT, INP-EI Purpan, Castanet-Tolosan, France

6 <sup>2</sup>Plant Pathology, Institute of Integrative Biology, Zurich (IBZ), ETH Zurich, LFW B16, 8092 Zurich, Switzerland

7 <sup>3</sup>Current address: School of Agriculture, Policy and Development, University of Reading, Whiteknights, Reading,

8 RG6 6AR United Kingdom

9 \* Corresponding author. Email: Laetitia.willocquet@inra.fr

10

## 11 Abstract

12 Plant disease emergences have dramatically increased recently as a result of global changes, especially with  
13 respect to trade, host genetic uniformity, and climate change. A better understanding of the conditions and  
14 processes determining epidemic outbreaks caused by the emergence of a new pathogen, or pathogen strain, is  
15 needed to develop strategies and inform decisions to manage emerging diseases. A polyetic process-based  
16 model is developed to analyse conditions of disease emergence. This model simulates polycyclic epidemics  
17 during successive growing seasons, the yield losses they cause, and the pathogen survival between growing  
18 seasons. This framework considers an immigrant strain coming into a system where a resident strain is already  
19 established. Outcomes are formulated in terms of probability of emergence, time to emergence, and yield loss,  
20 resulting from deterministic and stochastic simulations. An analytical solution to determine a threshold for  
21 emergence is also derived. Analyses focus on the effects of two fitness parameters on emergence: the relative  
22 rate of reproduction (speed of epidemics), and the relative rate of mortality (decay of population between  
23 seasons). Analyses revealed that stochasticity is a critical feature of disease emergence. The simulations  
24 suggests that: (1) emergence may require a series of independent immigration events before a successful

25 invasion takes place; (2) an explosion in the population size of the new pathogen (or strain) may be preceded by  
26 many successive growing seasons of cryptic presence following an immigration event, and; (3) survival between  
27 growing seasons is as important as reproduction during the growing season in determining disease emergence.

28

## 29 KEYWORDS

30 disease emergence, process-based model, pathogen fitness, polyetic epidemics, pathogen survival

31

## 32 1. INTRODUCTION

33 The emergence of disease in plant populations has important impacts on both agricultural production and  
34 natural ecosystems (Anderson *et al.*, 2004; Lucas, 2017). While emerging plant diseases threaten biodiversity  
35 and the entire range of services contributed by plants to the biosphere (Anderson *et al.*, 2004), the emergence  
36 of plant diseases constitutes an immediate threat to food security, from local to global scales, because of the  
37 losses in production, and also because losses to plant disease affect food access (economic or physical) and the  
38 quality of food (Savary *et al.* (2017). The literature provides growing evidence that plant disease emergences  
39 have dramatically increased recently, as a result of global changes in trade, host genetic uniformity, and climate  
40 (Anderson *et al.*, 2004; Fisher *et al.*, 2012; McDonald and Stukenbrock, 2016; Paine *et al.*, 2016).

41 A relatively recent example of emergence of new pathogen strains is the introduction into Europe of a  
42 strain carrying the A2 mating type of *Phytophthora infestans*, the causal agent of potato late blight  
43 (Zwankhuizen and Zadoks, 2002; Lucas, 2017). The emergence of this strain and its lineages, both resistant to  
44 metalaxyl and more aggressive, led to more diversified, sexually reproducing, pathogen populations, and  
45 increased disease intensity in Europe (Goodwin *et al.*, 1996). Stem rust of wheat is another example. Stem rust  
46 epidemics, which were common in the USA during the first half of the last century, became rare after the  
47 pathogen (*Puccinia graminis* f. sp. *tritici*) was controlled by combining the deployment of new resistance genes  
48 in wheat varieties with the eradication of barberry, which is the alternate host on which the pathogen

49 reproduces sexually (Roelfs, 1978; 1985). In 1998, new races of this pathogen (called Ug99) were detected in  
50 Uganda that were virulent against resistance genes present in wheat varieties widely grown in East Africa,  
51 leading to local but severe epidemics in the region (Singh *et al.*, 2015). International efforts to generate and  
52 deploy resistant varieties helped to limit impacts from races of these new lineages (Singh *et al.*, 2015), but the  
53 recent detection of stem rust in different parts of Europe is now threatening wheat production in this part of  
54 the world (Saunders *et al.*, 2019). A third and recent example of strain introduction is that of *Puccinia striiformis*  
55 f. sp. *tritici*, the causal agent of stripe (yellow) rust of wheat, into North-Western Europe in 2011 (de Vallavieille-  
56 Pope *et al.*, 2018) which caused serious epidemics.

57         An example of emergence of a new pathogen is *Pyricularia graminis-tritici*, the cause of wheat blast.  
58 The disease was restricted to South America until 2016, when the pathogen was accidentally introduced and  
59 caused a severe outbreak in South Asia (Ceresini *et al.*, 2018). Another example of new pathogen emergence is  
60 the Asian soybean rust, caused by *Phakopsora pachyrhizi*, which was introduced into South America at the  
61 beginning of this century and has since severely impacted soybean production on that continent (Lucas, 2017).  
62 Rhizomania is a virus disease of sugar beet that was first detected in the United Kingdom in 1987 and has since  
63 spread, resulting in increasing numbers of epidemics (Gilligan *et al.*, 2007). Other recent examples of disease  
64 emergence with very disastrous impacts on perennial crops include huanglongbing on citrus in the New World  
65 (Gottwald, 2010) and *Xylella fastidiosa* on olive trees in Southern Europe (Saponari *et al.*, 2019).

66         Disease emergence may be associated with changes in the environment, especially, human-made  
67 changes. A much-debated example is the case of fusarium head blight of wheat (wheat scab), which has been  
68 associated with the maize-wheat rotation, and with no-till practices (Zadoks and Schein, 1979; McMullen *et al.*,  
69 2012). Another example is that of false smut of rice, which has been associated with the cultivation of hybrid  
70 rice (Savary *et al.*, 2017). A third example of environmental change-driven emergence is that of *Sclerotium*  
71 *rolfsii*, a tropical pathogen on legumes (among many other hosts) becoming prevalent in the state of New York  
72 as a result of warming climate (S. Pethybridge, Personal Communication).

73           In their seminal article, Heesterbeek and Zadoks (1987) proposed a mathematical theory of pandemics,  
74   with three phases: zero-order, first-order, and second-order epidemics. This theory considers two groups of  
75   processes, the spatial spread of disease and the accumulation of disease cycles within and across crop cycles, to  
76   analyse pandemics. While the zero-order epidemic is field-bound and polycyclic, the first-order epidemic is  
77   area-bound and polycyclic, and the second-order is both continental and polyetic. The present article is a  
78   response, some thirty years later, to this article. Figure 1 represents a synthesis of processes which may be  
79   associated with disease emergence, organised in three paths. Path 1 is the invasion of a new pathogen into an  
80   ecosystem, through introduction, establishment, and spread. Path 1 is exemplified by the wheat blast epidemic  
81   in Bangladesh. Path 2 is the emergence of disease in response to environmental changes in an ecosystem,  
82   where environmental changes lead to disease intensification, further leading to disease spread within entire  
83   (agro)systems. Path 2 is illustrated by fusarium head blight of wheat or false smut of rice. Path 3 is the  
84   emergence of new strains through evolutionary processes. Path 3 is illustrated by wheat stem rust in Sub-  
85   Saharan Africa. Emergence paths may be combined. For instance, Paths 1 and 3 are combined in the potato late  
86   blight epidemic of the 1990s in Western Europe; Paths 1, 2, and 3 are combined in the emergence of stripe rust  
87   in Western Europe.

88           Similar to the emergence and re-emergence of infectious diseases in humans (Wilcox and Collwell,  
89   2005), the emergence of plant diseases entails the consideration of biocomplexity, i.e., of complex systems,  
90   where the biology of pathogens and hosts, their genetics, the changing environments – both natural and  
91   human-made, and the social and economic structures (including plant health management systems) interact.  
92   The present analysis does not address the biocomplexity of plant disease emergence as a whole, but rather  
93   focuses on a fragment of Figure 1, with emphasis on Paths 1 (emerging pathogens) and 3 (emerging strains).  
94   Elements of Path 2 (environmental change) are subsumed in the form of stochastic features of the modelling  
95   work.

Here we present a series of hypotheses underpinning the processes at play in disease emergence.

These hypotheses involve both demography (epidemiology) and population genetics as follows:

(1) from an epidemiological standpoint, emergence is a polyetic process, i.e., it is a process spanning several consecutive crop seasons (Zadoks, 1974; Zadoks and Schein, 1979; Heesterbeek and Zadoks, 1987);

(2) this polyetic process is inherently stochastic because it entails random and abrupt changes in the pathogen and host populations (Shaw, 1994). The process is also affected by random fluctuations in the environment (Gilligan and Van den Bosch, 2008);

(3) an important determinant of successful emergence is the diversity in the population from which the emerging pathogen originates. We assume the pathogen (or pathogen strain) to be sampled by chance in a large genetic pool. The more diverse this pool, the higher the likelihood of fit to a given biological (hosts) and physical setting (McDonald and Stukenbrock, 2016);

(4) pathogen migration (introduction) is often the primary mechanism associated with disease emergence (McDonald and Stukenbrock, 2016);

(5) the level of crop losses associated with epidemics constitutes a useful metric for the impact of disease emergence (Savary *et al.*, 2006; 2017; 2019).

A range of models have been developed to analyse the dynamics of epidemics or pathogen populations over multiple crop seasons. Leonard (1977) analysed the dynamics of plant pathogen genotypes over seasons to investigate plant pathogen evolution under the gene-for-gene hypothesis. Since then, several polyetic models have considered cycles of epidemic processes (disease transmission in the presence of the host) followed by survival processes (pathogen decay in the absence of the host). Several models have considered one pathogen genotype in order to address, e.g., thresholds for persistence according to epidemiological parameters (e.g., Gubbins *et al.*, 2000; Madden and Van den Bosch, 2002), whereby persistence corresponds to disease emergence caused by invasion. These models were expanded to consider two pathogen genotypes to analyse the evolutionary dynamics of pathogen populations (Van den Berg *et al.*, 2011; Hamelin *et al.*, 2011).

Comparatively fewer stochastic polyetic models have been developed, showing chaotic polyetic patterns (Shaw, 1994), or guiding management strategies (with a spatially explicit stochastic model of sugar beet rhizomania; Gilligan *et al.*, 2007). To our knowledge, no model has yet been developed which simultaneously accounts for polyetic processes, stochasticity, and the occurrence of several pathogen genotypes. Furthermore, none of the polyetic models reported so far explicitly accounts for the impact of disease on yield loss.

The objectives of this work were to: (1) design a modelling framework to better define the conditions determining disease emergence, (2) illustrate the use of the model by considering fitness components that characterize the growth of the pathogen population during the growing season and its survival between growing seasons and analysing their effects on disease emergence, and (3) draw some conclusions on properties associated with disease emergence.

130

## 2. MATERIALS AND METHODS

### 2.1 Model requirements

133

Model specifications were established to build a structure incorporating processes related to population genetics, epidemiology, and crop losses in order to analyse the conditions associated with disease emergence and its effects on yield. The model and the outcomes of the analysis apply equally well to: (i) the situation when a strain of a new pathogen is emerging on the background of another, resident pathogen population infecting the same host, and (ii) the situation where a new, immigrant strain of a pathogen emerges on the background of an already established population of the same pathogen. For the sake of simplicity, we will only refer to the second situation below. The following requirements for the model were identified:

- Because disease emergence takes place over several crop seasons (Gilligan and Van den Bosch, 2008), the model generates dynamics of epidemics and pathogen survival over successive cropping cycles, i.e.,



it encapsulates *polyetic processes* (Zadoks, 1974; Zadoks and Schein, 1979; Heesterbeek and Zadoks, 1987).

- Epidemics of many plant diseases entail secondary infections occurring during a crop season. The model therefore considers *polycyclic epidemics* within crop seasons.
- The model involves *different strains of a pathogen* in order to account for the evolutionary processes involved in disease emergence.
- Disease emergence often originates from the *migration* of a new pathogen (Fig. 1, Path 1), or of a new pathogen strain (Fig. 1, Path 3), into an agrosystem. The model therefore incorporates an immigration process.
- Modelling of the dynamics of *primary inoculum* with varying numbers of propagules, originating from preceding crop seasons and/or from immigration, and decaying over time, is a requirement, because (1) primary inoculum enables the initiation of seasonal epidemics, (2) a migrating pathogen strain enters the system as primary inoculum, and (3) primary inoculum also constitutes the link between two seasonal epidemics, and therefore provides the bridge needed to consider polyetic epidemics.
- *Crop losses* are an essential feature of epidemics in agroecosystems. The model therefore translates multi-seasonal, polyetic epidemics into their impact on crop performance as yield losses.

## 2.2 Model description

The system considered in the model is 1 m<sup>2</sup> of a crop, under the "mean-field" hypothesis: the system considered is surrounded by systems with the same features and dynamics. This 1 m<sup>2</sup>-system and its surroundings are repeated in successive crop seasons separated by off-seasons. During any crop season, this system and its neighbours are considered in a steady-state relationship. In particular, incoming and outgoing inoculum between these systems cancel each other out, so that the net inflow/outflow balance is null. The

166 model time step is one day, so as to accommodate processes which can have fast dynamics, such as polycyclic  
167 processes.

168         We consider a crop that is grown in regular cropping cycles (Supplementary Figure 1). Each cropping  
169 cycle consists of the period when the crop is present (crop season) and the period when the crop is absent (off-  
170 season). The crop season starts from crop planting and ends at harvest, and has two phases: the crop  
171 establishment phase and the crop growth phase (or, shortly, the growing season). The duration of each of the  
172 two phases (*CEP*, crop establishment period, and *CGP*, crop growth period) can vary depending on the crop,  
173 crop type (winter or spring crop) and location. Simulations start at crop planting, and are run for 30 cropping  
174 cycles.

175         The model considers one host plant genotype, e.g. a variety of a given crop, which can be infected by  
176 two strains of a given pathogen: a local (or resident) strain and an exogenous (or immigrant) strain. The local  
177 pathogen strain is present at the beginning of the simulation, while the exogenous pathogen strain is  
178 introduced into the system during the course of the simulation. The local population consists of strains that are  
179 already well adapted to local conditions. This local population is represented in the model by one local strain  
180 which has fixed demographic parameters. The exogenous population is established in a range of conditions  
181 (outside from the system), which may differ from the conditions of the considered system. This population is  
182 therefore more diverse, and generally less well adapted to the local conditions of the system. It thus consists of  
183 strains with a broader range of fitness attributes compared to the local population. This exogenous population  
184 is represented in the model by one strain with a fitness that can vary over cropping cycles. This variation  
185 reflects the hypothesis that the exogenous strain is less well adapted to the local conditions than the resident  
186 strain, and therefore is less well adapted to the environmental variations over cropping cycles. Each cropping  
187 cycle therefore involves two strains of the pathogen, the (fixed) local strain, and the (variable, random)  
188 exogenous strain.

Each cropping cycle involves several processes, which are represented as rates (Forrester, 1961; Savary and Willocquet, 2014) in Figure 2. These are the processes involved in the development of epidemics, including primary and secondary infections:  $RI$  (rate of infection); processes involved in the survival and decay of inoculum:  $Rdecay$  (rate of inoculum decay); and processes involved in yield losses incurred from disease:  $RL$  (rate of loss). These processes are next described in greater detail. The model variables and parameters are described in Table 1.

In each cropping cycle, the epidemic starts with primary infections ( $RPI$ ), which take place at the end of the crop establishment phase, as the crop growth phase starts. Primary infections have two origins. First, primary infections can originate from inoculum produced from epidemics which took place in previous crop seasons (polyetism), and second, primary infections can result from incoming inoculum (immigration from an exogenous population). In the beginning of the crop growth phase, the rate of primary infections for each strain,  $j = 1$  (local), or 2 (immigrant), is therefore written as:

$$RPI_j = convSP \times S_j + RM_j \quad (1)$$

Where  $convSP$  is the conversion of surviving inoculum into a rate of primary infections;  $S_j$  is the number of surviving propagules for each strain; and  $RM_j$  is the rate of infections originating from immigrant strains, referred to as the rate of immigration. The rate of primary infections,  $RPI_j$ , has the value given by Eq. (1) only on the first time step of each growing season and is set to zero at all other times.

An epidemic takes place as the injury level,  $i$ , increases according to a logistic curve (exponential increase of secondary infections, limited by the carrying capacity of the host crop) with a relative rate of growth,  $RRg$ . As the seasonal epidemic unfolds, interaction between strains takes place, in the form of competition towards host (crop) sites. This interaction between strains accounts for the maximum possible level of injury (carrying capacity) at a given time, considering all plant sites occupied by the different strains at this time. The rate of infection of each strain  $j$ , comprising primary and secondary infections, is therefore written as:

$$RI_j = [RRg_j \times i_j \times (1 - ((i_1 + i_2) / imax))] + RPI_j + starter_j \quad (2)$$

213 where  $RRg_j$  is the relative rate of injury increase for strain  $j$ ;  $i_j$  is the injury level of strain  $j$ ;  $i_1$  is the injury level  
 214 caused by the local strain;  $i_2$  is the injury level caused by the immigrant strain;  $imax$  is the carrying capacity of  
 215 injury, i.e., the maximum level of injury;  $RPI_j$  is the rate of primary infections associated with strain  $j$ ; and  $starter_j$   
 216 is the number of primary infections at the beginning of the multiple-cropping cycle simulation (this parameter is  
 217 non-zero only during the first time step of the cropping cycle 1).

218 At the end of a cropping cycle, the terminal injury level ( $i_j$ ) is converted into surviving inoculum,  $S_j$ , for  
 219 each of the two strains. The number of surviving propagules decreases over time according to a negative  
 220 exponential dynamics, at a speed proportional to a relative rate of decay ( $RRD_j$ ):

$$221 \quad Rdecay_j = RRD_j \times S_j, \quad (3)$$

222 where  $Rdecay_j$  is the rate of decay of surviving propagules of strain  $j$ ;  $RRD_j$  is the relative rate of decay of  
 223 surviving propagules of strain  $j$ , and  $S_j$  is the number of surviving propagules of strain  $j$ .

224 Injuries impair the physiological processes involved in crop growth and yield build-up, ultimately  
 225 leading to yield losses. The several possible damage mechanisms from injuries are represented in a very  
 226 simplified manner by a single rate of yield loss,  $RL$ , which increases proportionally to the running level of  
 227 combined injuries caused by both strains,  $i_1 + i_2$ :

$$228 \quad RL = RRL \times (i_1 + i_2) \times [1 - (YL / Ya)] \quad (4)$$

229 where  $RRL$  is the relative rate of yield loss;  $i_1$  and  $i_2$  are the injury levels from the local and immigrant strains,  
 230 respectively;  $YL$  is the yield loss, i.e. the yield reduction from a disease-free attainable yield; and  $Ya$  is the  
 231 attainable yield, i.e., the yield level in the absence of disease. At the end of each crop growth phase, yield loss is  
 232 reset to zero, so that the new cropping cycle starts without losses.

233

234

235

236 2.3 Model parameters and initial values

Initial values are zero for all state variables ( $i$ ,  $S$ , and  $YL$ ). Parameters dimensions and values are listed in Table 1. The durations of the crop establishment period ( $CEP$ ) and of the crop growth period ( $CGP$ ) are both set to 120 days, representing, for example, approximate durations for a winter wheat crop grown in a temperate region of the world. An epidemic of the local strain is initiated at the end of  $CEP$  in the first cropping cycle with an initial value ( $starter$ ; Table 1) for the local strain of  $0.01 \text{ day}^{-1}$ . The conversion of surviving propagules into a rate of primary infections ( $convSP$ ) is set to 0.01, meaning that for example 100 surviving entities are translated into 1 primary infection during the first time step of  $CGP$ . The carrying capacity for injury level,  $imax$ , is set to 100 in order to generate injury levels expressed as percent. In the same way,  $Ya$ , the attainable yield, is set to 100 in order to generate yield losses expressed as percent.

$RRL$  is set to 0.05, meaning that combined disease injury ( $i_1 + i_2$ ), when at low levels, entails an increase in yield loss at each time step which corresponds to 5% of the level of disease injury.  $RRg_1$  and  $RRD_1$  values are set to 0.07 and 0.01, respectively.

## 2.4 Model analyses: conditions of emergence of an immigrant strain

### 2.4.1 Framework of analyses

We consider a pathosystem with two pathogen strains: a local (resident), and an immigrant (exogenous) strain. The fitness of each of the two strains is represented by two essential components: the ability to reproduce during the growing season [represented by a relative, or intrinsic, rate of growth,  $RRg_j$  in Equation (2)] and the rate of population decay [represented by a relative, or intrinsic, rate of decay,  $RRD_j$  in Equation (3), Table 1], the latter characterizing the ability of a pathogen strain to survive in the absence of host plants. As a convention, the subscript  $j=1$  refers to a local strain and  $j=2$  refers to an immigrant strain. Fitter strains reproduce faster on the host during the growing season and decay more slowly over time.

We addressed the question of emergence of immigrant strains as follows. A given agroecosystem harbours a resident diversity of strains; however, all these strains are assumed to be equally adapted to the

261 considered agroecosystem – i.e. they have similar fitness. As a simplification, the entire population of resident  
262 strains in an agroecosystem is therefore represented by one strain, exhibiting two central values for  $RRg_1$  and  
263  $RRD_1$ . Because these local populations are assumed to be established and in a dynamic equilibrium, we further  
264 assume no variation over time for parameters  $RRg_1$  and  $RRD_1$ . In the absence of immigration, successive  
265 epidemics occur in the considered agroecosystem. These epidemics consist of overlapping disease cycles  
266 (polycyclic epidemics), and each epidemic results from the carry-over of inoculum from a previous epidemic  
267 that took place in the previous crop seasons. The resulting pattern of disease over successive crop seasons  
268 (polyetic process) in an agroecosystem thus results from the concatenation of successive (polycyclic) epidemics.

269 In order to investigate conditions for emergence, we consider an immigrant strain, which originates  
270 from a very large pool of possible strains. In a first (deterministic) regime, the fitness parameters of the  
271 immigrant strain,  $RRg_2$  and  $RRD_2$ , are assumed to be constant throughout the successive simulated cropping  
272 seasons. In a second (stochastic) regime, the fitness parameters of the immigrant strain are drawn at random  
273 from a normal distribution with central values  $RRg_2$  and  $RRD_2$ , and with variation about these values. This  
274 drawing is made at the beginning of each cropping cycle, and the values drawn are kept constant within each  
275 cropping cycle. This stochastic regime reflects the hypothesis of a strain which is not well adapted to the local  
276 environment, with a fitness that varies as environmental conditions vary over cropping cycles.

277 The execution of the model over a succession of 30 cropping cycles is referred to as a simulation. We  
278 investigated a scenario in which the immigrant strain is introduced once, at cropping cycle 10, at the beginning  
279 of the growing season. This way, the immigrant strain is introduced into a stabilized system where the local  
280 strain is already established. We used the simulation model (Section 2.2, Figure 2) to study two dynamic  
281 regimes: (i) a deterministic regime, in which  $RRg_2$  and  $RRD_2$  had fixed values during a given simulation (section  
282 2.4.2 below), and (ii) a stochastic regime, in which the values of either  $RRg_2$  or  $RRD_2$  or both were drawn from a  
283 normal (Gaussian) distribution at the beginning of each cropping cycle and kept at these values during each  
284 cropping cycle (section 2.4.3 below). We also derived approximate analytical expressions for the thresholds of

285 emergence of the immigrant strain by representing the simulation model as a discrete time map and  
286 investigating its linear stability (Section 2.4.4 below and Appendix A).

287         The outcomes of the analyses were synthesised according to three features characterising disease  
288 emergence of an immigrant strain and its consequences: the probability of emergence, the time to emergence,  
289 and the yield loss associated with the emergence. We consider that the immigrant strain has emerged if it  
290 exceeds the resident strain in terms of its AUDPC (area under disease progress curve, i.e., the accumulated  
291 injury incurred within a growing season) during at least three cropping cycles after its introduction. The  
292 probability of emergence,  $P_{\text{emerg}}$ , was estimated as the proportion of simulations that resulted in emergence. In  
293 each individual simulation that resulted in emergence, the time to emergence,  $T_{\text{emerg}}$ , was defined as the  
294 number of cropping cycles between the introduction of the immigrant strain and the first cropping cycle when  
295 the AUDPC of the immigrant strain exceeded that of the resident strain. To quantify yield loss in each  
296 simulation, we calculated the average yield losses caused by both the resident and the immigrant pathogen  
297 strains over the 30 cropping cycles.

298         The model was developed using the Stella software (STELLA Architect version 1.1.2) and subsequently  
299 translated to the Python programming language (version 3.4.3), where the bulk of the analysis was conducted.  
300 The system of Equations (1)-(4) was solved and analysed using Python packages numpy (version 1.13.3) and  
301 scipy (version 1.0.0), and the figures were produced using the Python package matplotlib (version 2.1.1). Parts  
302 of the analytical investigation were performed with Wolfram Mathematica (version 10.3 for Linux).

303

#### 304 2.4.2. Deterministic approach

305 We performed three sets of simulations in order to analyse the individual effects of  $RRg_2$ ,  $RRD_2$ , and the  
306 combined effects of  $RRg_2$  and  $RRD_2$  on disease emergence.

307         A first analysis was conducted to address conditions of emergence associated to  $RRg_2$ . In this first  
308 analysis, 100 simulations were run with  $RRg_2$  increasing from 0.06 to 0.12 day<sup>-1</sup> with a constant increment of

309  $RRg_2$  between simulations, while  $RRD_2$  was fixed ( $0.02 \text{ day}^{-1}$ ). The  $RRD_2$  value chosen corresponds to the  
310 hypothesis of an immigrant strain with a lower survival capacity than the resident strain ( $RRD_1 = 0.01 \text{ day}^{-1}$ ). In  
311 the second analysis, we assessed conditions of emergence according to  $RRD_2$ . Here, 100 simulations were run  
312 with  $RRD_2$  increasing from 0 to  $0.05 \text{ day}^{-1}$  with a constant increment between simulations, while  $RRg_2$  was fixed  
313 ( $0.1 \text{ day}^{-1}$ ). This  $RRg_2$  value corresponds to the hypothesis of an immigrant strain with a higher aggressiveness  
314 than the resident strain ( $RRg_1 = 0.07 \text{ day}^{-1}$ ). In a third analysis, both  $RRg_2$  and  $RRD_2$  were considered with respect  
315 to emergence.  $RRg_2$  and  $RRD_2$  were varied in the same ranges as in the first and second analyses over a total of  
316  $10^4$  simulations (100 x 100 runs).

317

#### 318 2.4.3. Stochastic approach

319 As in the deterministic approach, the individual effects of  $RRg_2$ ,  $RRD_2$ , and combined effects of  $RRg_2$  and  $RRD_2$   
320 were subsequently analysed.

321 To address conditions of emergence associated to  $RRg_2$ , 100 sets of simulations were executed with  
322 fixed  $RRg_2$  values ranging from  $0.06$  to  $0.12 \text{ day}^{-1}$ , with a constant increment. For each  $RRg_2$  value considered  
323 (i.e., for each set of simulations), 5000 stochastic runs were executed, within which the values of  $RRD_2$  were  
324 drawn at the beginning of each cropping cycle as random numbers from the normal (Gaussian) distribution with  
325 the mean  $0.02 \text{ day}^{-1}$  and the standard deviation  $0.007 \text{ day}^{-1}$ . These  $RRD_2$  values then remained constant during  
326 the whole cropping cycle until the beginning of the next growing season, when a new random value was  
327 chosen.

328 The second analysis was conducted in the same way as the first analysis, but focused on  $RRD_2$ : 100 sets  
329 of simulations were executed with fixed  $RRD_2$  values ranging from 0 to  $0.05 \text{ day}^{-1}$ , with a constant increment.  
330 For each  $RRD_2$  value considered, 5000 stochastic runs were executed, within which the values of  $RRg_2$  were  
331 drawn at the beginning of each cropping cycle as random numbers from the normal distribution with mean 0.1



332 day<sup>-1</sup> and standard deviation 0.035 day<sup>-1</sup>. These  $RRg_2$  values then remained constant during the whole cropping  
333 cycle.

334 In a third analysis, the values of both  $RRg_2$  and  $RRD_2$  were drawn from the normal distribution at the  
335 beginning of each cropping cycle with means ranging from 0.05 to 0.12 day<sup>-1</sup> for  $RRg_2$ , and ranging from 0 to  
336 0.04 for  $RRD_2$ , and with standard deviations constituting a constant proportion, 0.35, of the corresponding  
337 mean values. As in the previous analyses,  $RRg_1$  and  $RRD_1$  values remained constant within each cropping cycle.  
338 We ran 200 stochastic realizations for each point of the 100x100 grid of  $RRg_2$  x  $RRD_2$  values considered.

339

#### 340 2.4.4. Analytical approach

341 The overall fitness of the pathogen strain  $j$  is given by its polyetic (or multi-season) basic reproductive number  
342 (see Appendix A for the derivation):

$$343 R_{op,j} = convSP \times \exp[(RRg_j \times T_s) - (RRD_j \times T_{BS})] \quad (5)$$

344 where  $T_s$  is the crop growth duration ( $T_s = CGP$ ) and  $T_{BS}$  is the delay between two successive growing seasons.

345 The index "p" in  $R_{op}$  refers to "polyetic", in order to distinguish  $R_{op}$  from  $R_0$ , which usually refers to the "within  
346 season" basic reproductive number in the epidemiological literature (e.g., Zadoks and Schein, 1979; Anderson  
347 and May, 1986; Campbell and Madden, 1990). Biologically,  $R_{op,j}$  represents the number of units of crop injury  
348 appearing at the beginning of a given growing season following the introduction of a unit host injury in the  
349 beginning of the previous growing season.  $R_{op,j}$  incorporates both the ability of a strain  $j$  to multiply during crop  
350 growth and to survive between growing seasons. Hence, in the exponent of Eq. (5), the two components of  
351 pathogen fitness,  $RRg_j$  and  $RRD_j$ , are weighted by  $T_s$  and  $T_{BS}$ , respectively.

352 As in the previous analysis, we consider the situation when the local pathogen strain is viable when  
353 present alone: its reproduction during the growing season exceeds its losses between growing seasons, i.e.,  
354  $R_{op,1} > 1$ . In this case, the immigrant strain will emerge if its polyetic basic reproductive number exceeds the  
355 polyetic basic reproductive number of the local strain, i.e.,  $R_{op,2} > R_{op,1}$ . The emergence threshold corresponds to

356  $R_{0p,2} = R_{0p,1}$ . We solve this equation with respect to  $RRg_2$  and obtain the threshold value of  $RRg_2$  above which  
 357 emergence takes place:

$$358 \quad RRg_{2,thresh} = [(T_{BS} / T_S) \times (RRD_2 - RRD_1)] + RRg_1. \quad (6)$$

359 Similarly, the emergence threshold can be expressed in terms of  $RRD_2$ :

$$360 \quad RRD_{2,thresh} = [(T_S / T_{BS}) \times (RRg_2 - RRg_1)] + RRD_1. \quad (7)$$

361 Here, the immigrant strain emerges when its relative rate of decay is below the threshold, i.e.,  $RRD_2 < RRD_{2,thresh}$ .

362 Eqs. (6) and (7) were derived under the assumption that the saturation effects of the logistic growth are  
 363 negligible. This is justified at sufficiently low relative rates of growth for each of the strains  $RRg_j$ , and at short  
 364 enough CGP, so that the host tissue does not become a limiting factor for any of the two pathogen strains. Note  
 365 that the simulation model in Sec. 2.2 does not make this approximation. See Appendix A for more mathematical  
 366 details.

367

### 368 **3. RESULTS**

#### 369 **3.1 An example of dynamics of crop injuries and losses simulated with the stochastic approach**

370 Figure 3 displays examples of simulated dynamics using fitness parameters for the immigrant strain drawn from  
 371 a normal distribution at the beginning of each cropping cycle, the parameter values remaining fixed within a  
 372 given cropping cycle. Means of  $RRg_2$  and  $RRD_2$  are equal to the values used for the local strain (0.07 for  $RRg$  and  
 373 0.01 for  $RRD$ ) and their standard deviations are 0.03 and 0.003, respectively. The three top panels display injury  
 374 dynamics leading to (1) non-emergence of the immigrant strain (Fig. 3a), (2) co-occurrence of both strains  
 375 where the predominant strain varies over cropping cycles (Fig. 3b), and (3) rapid emergence of the immigrant  
 376 strain (Fig. 3c). Because parameters are fixed for the local strain, simulation leading to non-emergence (Fig. 3a)  
 377 shows an equilibrium state with a maximum level of injury which reaches a constant value starting from the 6<sup>th</sup>  
 378 cropping cycle. When both strains co-exist (Fig. 3b), the stochasticity of  $RRg_2$  and  $RRD_2$  produces a large  
 379 variation over cropping cycles in both injury level and the respective frequency of each strain. In the case of

380 rapid emergence of the immigrant strain (Fig. 3c), the immigrant strain very quickly overcomes the local strain,  
381 but displays a large variation in disease intensity over cropping cycles, because of stochasticity in  $RRg_2$  and  
382  $RRD_2$ .

383 Simulated yield loss, primary inoculum, and relative rates of growth and decay corresponding to the  
384 example of a rapidly emerging strain (Fig. 3c) are displayed in Figure 3d-f. Yield losses vary over time (Fig. 3d),  
385 with a pattern similar to that observed for injury dynamics (Fig. 3c). At the end of each crop season, the  
386 terminal disease injury from each strain is proportionally converted to surviving propagules. The number of  
387 surviving propagules then decays exponentially over time and constitutes the primary inoculum for the  
388 subsequent growing season (Fig. 3e). This primary inoculum translates into primary infections at the beginning  
389 of each crop growth phase (Fig. 3e). Relative rates of growth and of decay of the immigrant strain vary over  
390 cropping cycles, while remaining constant within each cropping cycle (Fig. 3e). These stochastic values of  $RRg_2$   
391 and  $RRD_2$  are driving the dynamics of injury (Fig. 3c) and of primary inoculum (Fig. 3e) over cropping cycles.

392

### 393 3.2 Individual effects of the relative rates of epidemic growth and inoculum decay on disease emergence

394 When we consider  $RRg_2$  variation in the deterministic regime,  $P_{\text{emerg}}$  rises suddenly from zero to one as the  $RRg_2$   
395 value is increased (Fig. 4a, blue curve). The reason is that the immigrant strain can only emerge if it is able to  
396 grow fast enough during the growing season. More fit immigrants, when they emerge (i.e., when  $P_{\text{emerg}} = 1$  in  
397 Fig. 4a), emerge more rapidly as  $RRg_2$  increases:  $T_{\text{emerg}}$  decreases monotonically as we increase the immigrant  
398 strain's fitness by increasing  $RRg_2$  (Fig. 4c, blue curve). As we increase  $RRg_2$ , the amount of disease caused by  
399 the immigrant strain increases and so does the average yield losses incurred by both the resident and the  
400 immigrant pathogen strains (Fig. 4e, blue curve). Below the emergence threshold (Fig. 4a,  $RRg_2$  values for which  
401  $P_{\text{emerg}} = 0$ ), the immigrant strain is absent, therefore the yield losses are only incurred by the resident strain, and  
402 are not affected by  $RRg_2$ . The analytical approach yields a threshold for emergence of  $RRg_2$  at  $0.09 \text{ day}^{-1}$  (Fig. 4a,

grey line), that is, slightly smaller than the threshold derived from the deterministic approach (Fig. 4a, blue line).

When we include stochasticity, the transition between parameter areas of “no emergence” and “emergence” is now gradual:  $P_{\text{emerg}}$  increases continuously as  $RRg_2$  values of the immigrant strain are increased (red curve in Fig. 4a). Time to emergence also exhibits a different pattern in the stochastic regime. Emergence starts at  $RRg_2$  values much smaller than in the deterministic regime (Fig. 4c, red curve) with relatively small values of  $T_{\text{emerg}}$  (about 2 cropping cycles), then  $T_{\text{emerg}}$  increases gradually, plateaus at about four cropping cycles, and then declines with a curve close to, but above, that generated from the deterministic approach. Yield losses show a similar pattern in the deterministic and stochastic regimes (compare red and blue curves in Fig. 4e), although yield losses are somewhat lower in the stochastic regime than in the deterministic regime.

The effect of  $RRD_2$  on emergence characteristics (Fig. 4b, d, f) mirrors the effect of  $RRg_2$  (Fig. 4a, c, e), because as fitness of the immigrant strain increases with  $RRg_2$ , it decreases with  $RRD_2$ . Under the deterministic regime,  $P_{\text{emerg}}$  drops abruptly from one to zero as  $RRD_2$  is increased (Fig. 4b, blue curve): the immigrant strain cannot emerge if its population decays too fast between growing seasons. In the same way, less fit immigrants emerge more slowly, when they do emerge:  $T_{\text{emerg}}$  increases monotonically as we reduce the immigrant strain's fitness as  $RRD_2$  increases (Fig. 4d, blue line). Average yield losses decrease as  $RRD_2$  increases (Fig. 4f), and remain stable when  $RRD_2$  values are above the threshold for emergence. The threshold for emergence generated from the analytical approach,  $RRD_2 = 0.025$ , is slightly lower than the threshold generated from the deterministic approach (Fig. 4b).

When we include stochasticity,  $P_{\text{emerg}}$  diminishes continuously as  $RRD_2$  is increased; time to emergence,  $T_{\text{emerg}}$ , increases initially (with values slightly larger than those obtained from the deterministic approach), reaches a maximum around the emergence threshold and gradually declines to small values. Yield losses show a qualitatively similar pattern in the deterministic and stochastic regimes. As when investigating the effect of  $RRg_2$ , yield losses are lower in the stochastic regime than in the deterministic regime (Fig. 4f).

427

### 428 3.3 Combined effects of the relative rates of epidemic growth and inoculum decay on disease emergence

429 When using the deterministic approach, emergence and no emergence domains are clearly separated by a  
430 straight line (Fig. 5a). This line reflects the abrupt transition from emergence to no emergence, as illustrated in  
431 Figs. 4a and 4d. The domain of emergence corresponds to pairs of values of  $RRg_2$  and  $RRD_2$  below which  
432 emergence takes place: for a given value of  $RRg_2$ , emergence will occur within a range of values of  $RRD_2$  below a  
433 given threshold. The dashed lines in Fig. 5a display the outcomes for fitness values used which correspond to  
434 that of the resident strain. At  $RRg_2 = RRg_1 = 0.07$ , emergence occurs for  $RRD_2$  values slightly smaller than  $RRD_1$ .  
435 Similarly, at  $RRD_2 = RRD_1 = 0.01$ , emergence occurs for  $RRg_2$  values that are slightly larger than  $RRg_1$ . The solid  
436 grey line represents the analytical expression for the emergence threshold in terms of  $RRD_2$ , according to  
437 Equation (7). That is, a line with slope  $(T_s / T_{BS})$  which equals 0.5 in our case, and an ordinate at origin of -0.025.

438 When stochasticity is included in the model, the transition between parameter domains of emergence  
439 and no emergence becomes gradual (Fig. 5b). This gradual transition is a generalisation of the gradual change in  
440 emergence probability according to  $RRg_2$  and  $RRD_2$  illustrated in Figs. 4a and 4d, respectively. As in Fig. 5a, the  
441 dashed black line represents the fitness parameters of the resident strain. In the same way, the parameter  
442 region explored in Fig. 4a and 4b is materialised with the yellow lines which refer to  $RRg_2 = 0.1$  and  $RRD_2 = 0.02$ .  
443 The most important effect of stochasticity is that emergence occurs at ranges of parameters that are below the  
444 emergence threshold, where no emergence was possible according to the deterministic approach (within the  
445 white area in Fig. 5a). Even if the immigrant strain is on average less fit than the local strain (i.e., in terms of its  
446 average fitness components  $RRg_2 < RRg_1$ ,  $RRD_2 > RRD_1$ ), there is still a non-zero probability for its emergence.  
447 This scenario corresponds to the region in Fig. 5b above the black horizontal line and to the left from the black  
448 vertical line.

449

450

## 451 4. DISCUSSION

### 452 4.1 Key findings

453 The modelling framework which was designed in this work enables analysing the conditions underlying disease  
454 emergence. According to our pre-set specifications, the model includes polyetism, stochasticity, and yield loss.  
455 The analyses conducted here allow identifying important features associated with disease emergence.

456 A major finding is that stochasticity can be an important boundary condition for disease emergence.  
457 Emergence reflects important changes in the status of a system (here crop health) that can be caused by rare  
458 events (e.g., Paini *et al.*, 2016), associated with small sizes of immigrant or mutant subpopulations that initiate  
459 the process, and by polyetic processes that lead to significant reductions in population size between growing  
460 seasons (Shaw, 1994). Stochasticity associated with genetic factors such as bottlenecks and genetic drift is  
461 known to play an important role in the evolution of a host-pathogen interaction (McDonald, 2004).  
462 Stochasticity can also be introduced by environmental factors such as climatic conditions, which can  
463 differentially affect the fitness of strains in pathogen populations (e.g., Gilligan and Van den Bosch, 2008). In the  
464 present study, we focused on the latter, environmentally-induced, stochasticity.

465 Previous modelling analyses considered polyetic processes (e.g., Gubbins *et al.*, 2000; Madden and Van  
466 den Bosch, 2002) under a deterministic framework, leading to identification of thresholds for persistence.  
467 Results generated from the stochastic approach in this work produce a different outcome, in showing that (1)  
468 even when an immigrant strain is drawn from a population which is, on average, less fit than the local strain,  
469 the immigrant strain may nevertheless emerge due to stochasticity; and conversely (2) even when an immigrant  
470 strain is drawn from a population which is, on average, more fit than the local strain, the immigrant strain will  
471 not necessarily emerge, and may face extinction. An important implication of this finding is that emergence  
472 may require a series of independent immigration events involving new pathogen strains before a successful  
473 invasion takes place. In some cases, a pathogen that appears to have suddenly emerged over the course of only  
474 1-2 cropping cycles may have been present at a low level for decades before the proper conditions (e.g.

475 conducive weather conditions) occurred to enable an explosion in population size and an observed  
476 "emergence". This has important implications to guide future research, both in terms of modelling and  
477 experimentation, and potentially to inform policy on emerging diseases.

478 Another important finding from our analyses is that survival between growing seasons is as important  
479 for emergence as the pathogen reproduction during the growing season. Pathogens with limited saprophytic  
480 abilities and lacking durable survival structures such as chlamydospores, sclerotia or oospores are expected to  
481 undergo large bottlenecks between host growing seasons that will purge genetic diversity and increase the  
482 probability that less well-adapted immigrants occurring at lower frequencies will go extinct between growing  
483 seasons. Conversely, pathogens that compete well as saprophytes and/or produce long-lived survival structures  
484 will maintain high effective population sizes that sustain high levels of genetic diversity across growing seasons,  
485 enabling persistence of immigrants and novel mutants for long periods of time, even if they are less well  
486 adapted, and increasing the probability that these immigrants, mutants or recombinants can make a successful  
487 invasion. Although the importance of the survival phase has been recognized in earlier work (e.g., Heesterbeek  
488 and Zadoks, 1987; Gubbins *et al.*, 2000; Madden and Van den Bosch, 2002; Hamelin *et al.*, 2011), survival has  
489 often been overlooked by plant pathologists. Conversely,  $RR_g$  can be seen as the apparent infection rate of Van  
490 der Plank ( $r_L$ ; Campbell and Madden, 1990), for which ranges have been measured from disease progress curves  
491 in many instances. The  $RR_g$  ranges explored in our analyses (0.05 to 0.12) fit well within ranges measured for  
492 epidemics of annual crop diseases (Kranz, 2003).

#### 493 494 4.2 Comparing outcomes from the deterministic, analytical, and stochastic approaches

495 There is a good agreement between the analytical emergence thresholds (Eq. (6) and (7)) and the numerical  
496 thresholds in the deterministic regime, although the threshold for emergence with respect to  $RR_g$  is slightly  
497 lower when derived from the analytical approach as compared to the deterministic approach (Figs. 4, 5a), while  
498 the opposite pattern is obtained for  $RR_D$  (Figs. 4b, 5a). This difference can be explained by the limited duration

499 (30 cropping cycles) of the numerical simulations. In some cases, the immigrant strain would be able to emerge,  
500 but this would require more than 30 cropping cycles. On the contrary, the analytical threshold does not restrict  
501 the number of cropping cycles and therefore generates thresholds for emergence that can occur over an infinite  
502 time span. This explanation was confirmed by performing additional simulations conducted using the same  
503 design that generated Figure 4, but including many additional cropping cycles (200). In that case, the agreement  
504 between the two thresholds (from deterministic simulations and from analytical expressions) was perfect. In  
505 future analyses using this framework, the threshold values to consider (from deterministic or analytical  
506 approaches) will depend on the modelling objectives and the applications under consideration.

507         When investigating the probability of emergence, fitness thresholds are derived from the deterministic  
508 approach, while such thresholds do not materialize in the stochastic approach because the emergence  
509 probability can take values between 0 and 1. The stochastic approach allows a strain with a fitness ( $RRg$  or  $RRD$ )  
510 mean value below the deterministic emergence threshold (i.e., lower values for  $RRg$  and larger values for  $RRD$ )  
511 to emerge, with a probability which progressively declines as the mean fitness value moves away from the  
512 threshold.

513         The time to emergence progressively declines as the fitness values increase in the deterministic  
514 approach because it requires progressively less time for the immigrant strain to outcompete the resident strain.  
515 Under the stochastic regime, a different pattern is exhibited, with the time for emergence increasing, reaching a  
516 maximum, and eventually declining as the fitness values increase. This pattern can be interpreted as follows: at  
517 low average fitness, the only way to achieve emergence in the stochastic regime is when high fitness values  
518 from the tail of the distribution are drawn over several consecutive growing seasons, representing particularly  
519 "lucky" realizations. There is a small proportion of such realizations (reflected by the small probability of  
520 emergence), as they correspond to quite rare events, but when they do happen, emergence occurs relatively  
521 fast. In contrast, at higher average fitness, there can be many other paths to emergence including those  
522 realizations in which high fitness values appear sporadically, not necessarily in several consecutive seasons,



523 leading to slower emergence on average. Thus, the two competing effects, (i) longer emergence due to reduced  
524 mean fitness of the immigrant strain in the range of high fitness values and (ii) the preferential emergence of  
525 only "lucky", "fast-emerging" realizations in the range of low fitness values, lead to emergence time reaching a  
526 maximum in the stochastic regime.

527 Yield losses derived from the stochastic regime are lower than yield losses derived from the  
528 deterministic approach (Figs. 4e, f). This difference can be seen as the consequence of differences observed  
529 between these regimes both in terms of the probability of emergence (Figs. 4a, b) and the time to emergence  
530 (Figs 4c, d): above the deterministic threshold, there are cases where disease does not emerge, or where time  
531 to emergence is delayed in the stochastic approach, and therefore yield losses are not as high as in the  
532 deterministic approach.

533

#### 534 4.3 Further questions to address on disease emergence

535 Our analyses provide a series of elements to better understand how disease emerges. The model structure  
536 allows addressing other important questions on disease emergence, such as:

537 - the effect of primary infection patterns on emergence: in the analyses we conducted, we considered only one  
538 type of primary infection, as a single immigration event occurring at a single point in time. The model allows the  
539 consideration of other patterns, including varying size of immigrant inoculum, or repeating inflows of immigrant  
540 strains over several cropping seasons (instead of during only one cropping season).

541 - the immigration rate simply considers the entry into the system of a new strain, with no specific hypothesis  
542 attached to the origin of this strain. The model also allows consideration of other potential sources of new  
543 strains, including recombinants or mutants, which could originate from inside or outside of the zone of  
544 emergence.

545 - the model can also include adaptation of the pathogen population, for example by varying  $RRg$  and  $RRD$  over  
546 time, or draw new parameters at the start of each cropping cycle according to the parameter values of the  
547 preceding cropping cycle (Figure 1, path 3).  
548 - the effect of variation of  $RRg$  within the growing season can be analysed in order to mimic the effect of  
549 weather (e.g. warmer winters or drier summers) on epidemics and emergence (Figure 1, Path 2).  
550 - the analyses were conducted with a relatively limited number of cropping cycles of simulation. This was  
551 appropriate because a large amount of inoculum was used in the simulations. When considering a low rate of  
552 immigration within a stochastic regime, much longer time frames may be needed to detect emergence.  
553 - the effect of climate change on disease emergence can be addressed with this model by incorporating a  
554 directional change in the mean and/or the standard deviation of some fitness parameters (e.g.,  $RRg$  and  $RRD$ )  
555 over successive cropping cycles.

556

557       Plant disease emergence is a complex phenomenon, with many system- and context-specific variants.  
558 This work addresses the phenomenon in a simplified manner in order to derive some of its main features.  
559 While this work needs to be continued, we hope that the present analysis provides a useful step towards  
560 implementing more effective policies to prevent or delay plant disease emergence.

561

## 562 **DATA AVAILABILITY STATEMENT**

563 The data that support the findings of this study are available from the corresponding author upon reasonable  
564 request.

565   **REFERENCES**

- 566   Anderson, P. K., Cunningham, A.A., Patel, N.G., Morales, F.J., Epstein, P.R. and Daszak, P. (2004) Emerging  
567       infectious diseases of plants: pathogen pollution, climate change and agrotechnology drivers. *Trends in*  
568       *Ecology and Evolution*, 19, 535-544.
- 569   Anderson, R.M. and May, R.M. (1986) The invasion, persistence and spread of infectious diseases within animal  
570       and plant communities. *Philosophical Transactions of the Royal Society of London B, Biological Sciences*, 314,  
571       533-570.
- 572   Campbell, C.L. and Madden, L.V. (1990) *Introduction to Plant Disease Epidemiology*. New York, USA: John Wiley  
573       and Sons.
- 574   Ceresini, P.C., Castroagudín, V.L., Rodrigues, F.Á., Rios, J.A., Aucique-Pérez, C.E., Moreira, S.I. *et al.* (2018) Wheat  
575       blast: past, present, and future. *Annual Review of Phytopathology*, 56, 427-456.
- 576   Fisher, M.C., Henk, D.A., Briggs, C.J., Brownstein, J.S., Madoff, L.C., McCraw, S.L. and Gurr, S.J. (2012) Emerging  
577       fungal threats to animal, plant and ecosystem health. *Nature*, 484, 186-194.
- 578   Forrester, J.W. (1961) *Industrial Dynamics*. Cambridge, Mass., USA: M.I.T. Press.
- 579   Gilligan, C.A. and Van den Bosch, F. (2008) Epidemiological models for invasion and persistence of pathogens.  
580       *Annual Review of Phytopathology*, 46, 385-418.
- 581   Gilligan, C.A., Truscott, J.E. and Stacey, A.J. (2007) Impact of scale on the effectiveness of disease control  
582       strategies for epidemics with cryptic infection in a dynamical landscape: an example for a crop disease.  
583       *Journal of the Royal Society Interface*, 4, 925-934.
- 584   Goodwin, S.B., Sujkowski, L.S. and Fry, W.E. (1996) Widespread distribution and probable origin of resistance to  
585       metalaxyl in clonal genotypes of *Phytophthora infestans* in the United States and western Canada.  
586       *Phytopathology*, 86, 793-799.
- 587   Gottwald, T.R. (2010) Current Epidemiological Understanding of Citrus Huanglongbing. *Annual Review of*  
588       *Phytopathology*, 48, 119-39.

589 Gubbins, S., Gilligan, C.A. and Kleczkowski, A. (2000) Population dynamics of plant–parasite interactions:  
 590 thresholds for invasion. *Theoretical Population Biology*, 57, 219-233.

591 Hamelin, F.M., Castel, M., Poggi, S. and D. and Mailleret, L. (2011) Seasonality and the evolutionary divergence  
 592 of plant parasites. *Ecology*, 92, 2159-2166.

593 Heesterbeek, J.A.P. and Zadoks, J. C. (1987) Modelling pandemics of quarantine pests and diseases: problems  
 594 and perspectives. *Crop Protection*, 6, 211-221.

595 Kranz, J. (2003) *Comparative Epidemiology of Plant Diseases*. Berlin, Germany: Springer.

596 Leonard, K.J. (1977) Selection pressures and plant pathogens. *Annals of the New York Academy of Sciences*, 287,  
 597 207-222.

598 Lucas, J.A. (2017) Fungi, food crops, and biosecurity: advances and challenges. In: *Advances in Food Security*  
 599 *and Sustainability*. Amsterdam, The Netherlands: Elsevier, pp. 1-40.

600 Madden, L.V., Van Den Bosch, F. (2002) A population-dynamics approach to assess the threat of plant pathogens  
 601 as biological weapons against annual Crops. *BioScience*, 52, 65-74.

602 McDonald, B.A. (2004) *Population Genetics of Plant Pathogens*. APSnet Education Center. The Plant Health  
 603 Instructor. doi:10.1094/PHI-A-2004-0524-01

604 McDonald, B.A., Stukenbrock, E. H. (2016) Rapid emergence of pathogens in agro-ecosystems: global threats to  
 605 agricultural sustainability and food security. *Philosophical Transactions of the Royal Society of London B*,  
 606 *Biological Sciences*, 371, 20160026.

607 McMullen, M., Bergstrom, G., De Wolf, E., Dill-Macky, R., Hershman, D., Shaner, G. and Van Sanford, D. (2012) A  
 608 unified effort to fight an enemy of wheat and barley: Fusarium head blight. *Plant Disease*, 96, 1712-1728.

609 Paini, D.R., Sheppard, A.W., Cook, D.C., De Barro, P.J., Worner, S.P. and Thomas, M.B. (2016) Global threat to  
 610 agriculture from invasive species. *Proceedings of the National Academy of Sciences*, 113, 7575-7579.

611 Roelfs, A.P. (1978) *Estimated losses caused by rust in small grain cereals in the United States—1918–76*. US  
 612 Department of Agriculture, Miscellaneous Publications, 1363, 85 pp.

613 Roelfs, A.P. (1985) Wheat and Rye stem rust. In: Roelfs, A.P. Bushnell, W.R. (Eds.), *The Cereal Rusts. Volume II:*  
614 *Diseases Distribution, Epidemiology, and Control*. Orlando, USA: Academic Press.

615 Saponari, M., Giampetruzzi, A., Loconsole, G., Boscia and D., Saldarelli, P. (2019) *Xylella fastidiosa* in olive in  
616 Apulia: where we stand. *Phytopathology*, 109, 175-186.

617 Savary, S. and Willocquet, L. (2014) *Simulation Modeling in Botanical Epidemiology and Crop Loss analysis*.  
618 APSnet Education Center. The Plant Health Instructor. doi: 10.1094/PHI-A-2014-0314-01.

619 Savary, S., Begaglio, S., Willocquet, L., Gustafson, D., Mason, D'Croz D., Sparks, A., et al. (2017) Crop health and  
620 its global impacts on the components of food security. *Food Security*, 9, 311-327.

621 Savary, S., Willocquet, L., Pethybridge, S.J., Esker, P., McRoberts, N. and Nelson, A. (2019) The global burden of  
622 pathogens and pests on major food crops. *Nature Ecology & Evolution*, 3, 430-439.

623 Savary, S., Teng, P.S., Willocquet, L. and Nutter, F.W. Jr. (2006) Quantification and modeling of crop losses: a  
624 review of purposes. *Annual Review of Phytopathology*, 44, 89-112.

625 Saunders, D.G., Pretorius, Z.A. and Hovmøller, M.S. (2019) Tackling the re-emergence of wheat stem rust in  
626 Western Europe. *Communications Biology*, 2, 51.

627 Shaw, M.W. (1994) Seasonally induced chaotic dynamics and their implications in models of plant disease. *Plant*  
628 *Pathology*, 43, 790-801.

629 Singh, R. P., Hodson, D.P., Jin, Y., Lagudah, E.S., Ayliffe, M.A., Bhavani, S., et al. (2015) Emergence and spread of  
630 new races of wheat stem rust fungus: continued threat to food security and prospects of genetic control.  
631 *Phytopathology*, 105, 872-884.

632 de Vallavieille-Pope, C., Bahri, B., Leconte, M., Zurfluh, O., Belaid, Y., Maghrebi, E., et al. (2018) Thermal  
633 generalist behaviour of invasive *Puccinia striiformis* f. sp. *tritici* strains under current and future climate  
634 conditions. *Plant Pathology*, 67, 1307-1320.

635 Van den Berg, F., Bacaer, N., Metz, J.A., Lannou, C. and van den Bosch, F. (2011) Periodic host absence can select  
636 for higher or lower parasite transmission rates. *Evolution Ecology*, 25, 121-137.

637 Wilcox, B.A. and Colwell, R.R. (2005) Emerging and reemerging infectious diseases: biocomplexity as an  
638 interdisciplinary paradigm. *EcoHealth*, 2, 244-257.

639 Zadoks, J.C. (1974) The role of epidemiology in modern phytopathology. *Phytopathology* 64, 918-923.

640 Zadoks, J.C. and Schein, R.D. (1979) *Epidemiology and Plant Disease Management*. New York, USA: Oxford  
641 University Press.

642 Zwankhuizen, M.J. and Zadoks, J.C. (2002) *Phytophthora infestans*'s 10-year truce with Holland: a long-term  
643 analysis of potato late-blight epidemics in the Netherlands. *Plant Pathology*, 51, 413-423.

644

## 645 Appendix A

646

647 Assuming that the dynamics of the two pathogen strains are independent of each other, the simulation model  
648 described in Sec. 2.2 [Eqs. (1)-(4)] can be summarised in a single equation:

$$649 \quad i_{j,t+1} = R(i_{j,t}) i_{j,t} \quad (A1)$$

650 Where

$$651 \quad R(i_{j,t}) = \frac{imax \times convSP \times \exp[(RRg_j \times T_s) - (RRD_j \times T_{BS})]}{imax + i_{j,t} [\exp(RRg_j \times T_s) - 1]} \quad (A2)$$

652 Here,  $i_{j,t}$  is the injury caused by the pathogen strain  $j$  ( $j=1$  for the local strain, and  $j=2$  for the immigrant strain) at  
653 the very beginning of the growing season  $t$ , where  $t$  is the index that runs through successive cropping cycles  
654 (i.e.,  $t = 1, 2, 3, \dots$ ).  $T_s$  is the duration of the growing season (= CGP) and  $T_{BS}$  is the duration between two successive  
655 growing seasons. Eq. (A1) is a map that relates the injury at the beginning of growing season  $t+1$ ,  $i_{j,t+1}$ , to the  
656 injury at the beginning of the previous growing season  $t$ ,  $i_{j,t}$ , representing a nonlinear generalisation of the  
657 classical geometric growth model. The map Eq. (A1) has two fixed points:

$$658 \quad i_{j,FP1} = 0; i_{j,FP2} = \frac{imax \times convSP \times \exp[(RRg_j \times T_s) - (RRD_j \times T_{BS})] - imax}{\exp(RRg_j \times T_s) - 1} \quad (A3)$$

659 The fixed points determine the long-term outcomes of the dynamics: eventually the strain  $j$  either dies out  
660 (FP1) or reaches the stable equilibrium (FP2). The equilibrium occurs due to a balance between pathogen  
661 reproduction during the growing season and its decay between growing seasons: the number of newly  
662 produced pathogen individuals during the growing season compensates the number of individuals lost during  
663 the preceding between-growing season phase. Which of the two fixed points is achieved in the long run, is  
664 determined by the growth rate of the map Eq. (A1) linearised in the vicinity of FP1:

$$665 \quad R_{op,j} = convSP \times \exp[(RRg_j \times T_s) - (RRD_j \times T_{BS})] \quad (A4)$$

666 where  $R_{0p,j}$  quantifies the reproductive fitness of the strain  $j$  and corresponds here to the polyetic basic  
667 reproductive number. Usually, the basic reproductive number is defined as the number of host individuals that  
668 are infected by a single infected host introduced into a fully susceptible host population (e.g., Zadoks and  
669 Schein, 1979; Anderson and May, 1986; Campbell and Madden, 1990). Adapted to our context, the biological  
670 meaning of  $R_{0p,j}$  is the number of units of crop injury appearing at the beginning of growing season  $t + 1$   
671 following the introduction of a unit host injury at the beginning of the previous growing season  $t$ . If each of the  
672 two strains is viable when present alone, i.e.,  $R_{0p,j} > 1$ , the strain that has a higher basic reproductive number  
673 eventually outcompetes the other strain. Consequently, the immigrant strain emerges if it has a higher polyetic  
674 basic reproductive number, i.e.,  $R_{0p,2} > R_{0p,1}$ . The emergence threshold is given by the equality of the two  
675 polyetic basic reproductive numbers:  $R_{0p,2} = R_{0p,1}$ . We solve this equation with respect to  $RRg_2$ , using Eq. (A4), to  
676 obtain the threshold value of  $RRg_2$  above which the emergence takes place:

$$677 \quad RRg_{2,thresh} = (T_{BS} / T_S) \times (RRd_2 - RRd_1) + RRg_1 \quad (A5)$$

678 Similarly, the emergence threshold can be expressed in terms of  $RRD_2$ :

$$679 \quad RRD_{2,thresh} = (T_S / T_{BS}) \times (RRg_2 - RRg_1) + RRD_1 \quad (A6)$$

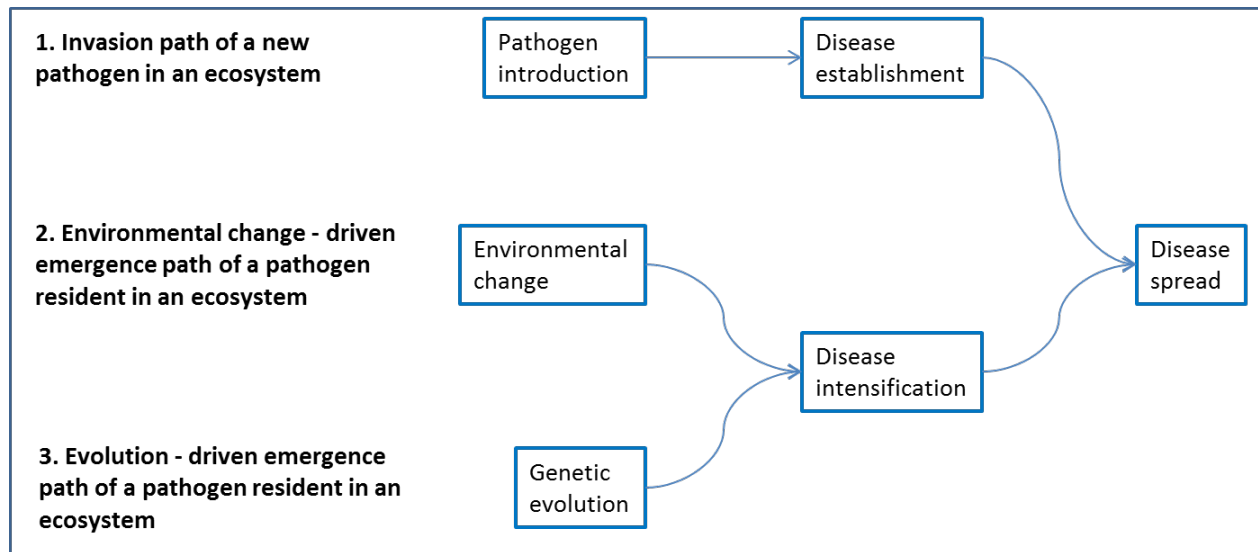
680 Here, the immigrant strain emerges when its relative decay rate is below the threshold, i.e.,  $RRD_2 < RRD_{2,thresh}$ .  
681 When only one pathogen strain is present,  $R_{0p,j}$  given by Eq. (A4) determines without any approximation which  
682 of the two fixed points in Eq. (A3) will be achieved according to the map in Eq. (A1). However, when both  
683 pathogen strains are present, Eq. (A4) and Eqs. (A5), (A6) derived from it, give only approximate expressions for  
684 emergence thresholds, under the assumption that the saturation effects of the logistic growth are negligible.  
685 Nevertheless,  $i_{j,FP2}$  in Eq. (A3) provides an exact expression for the final, equilibrium level of injury due to the  
686 pathogen strain that wins the competition.

687

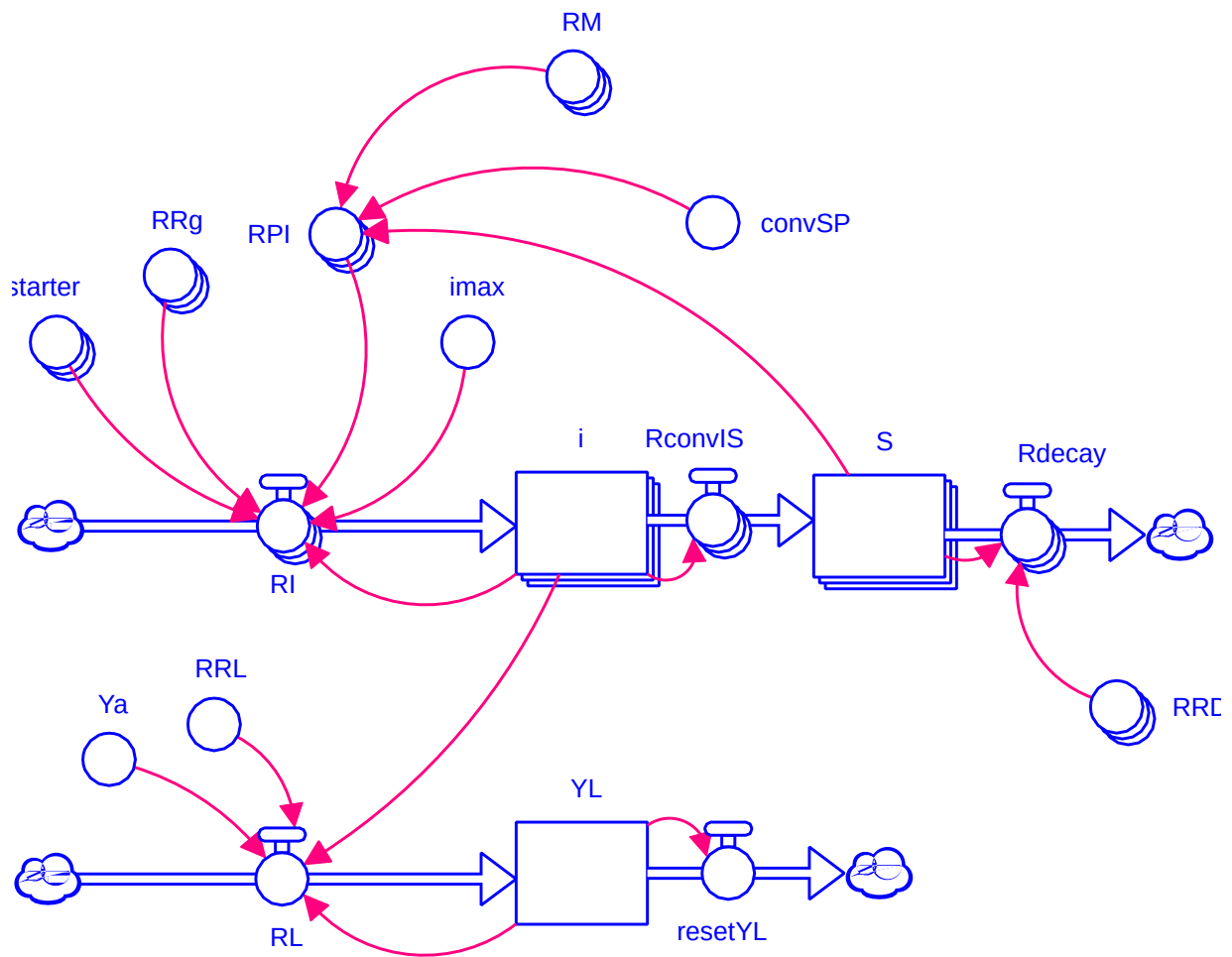


**Table 1.** Description of the model variables

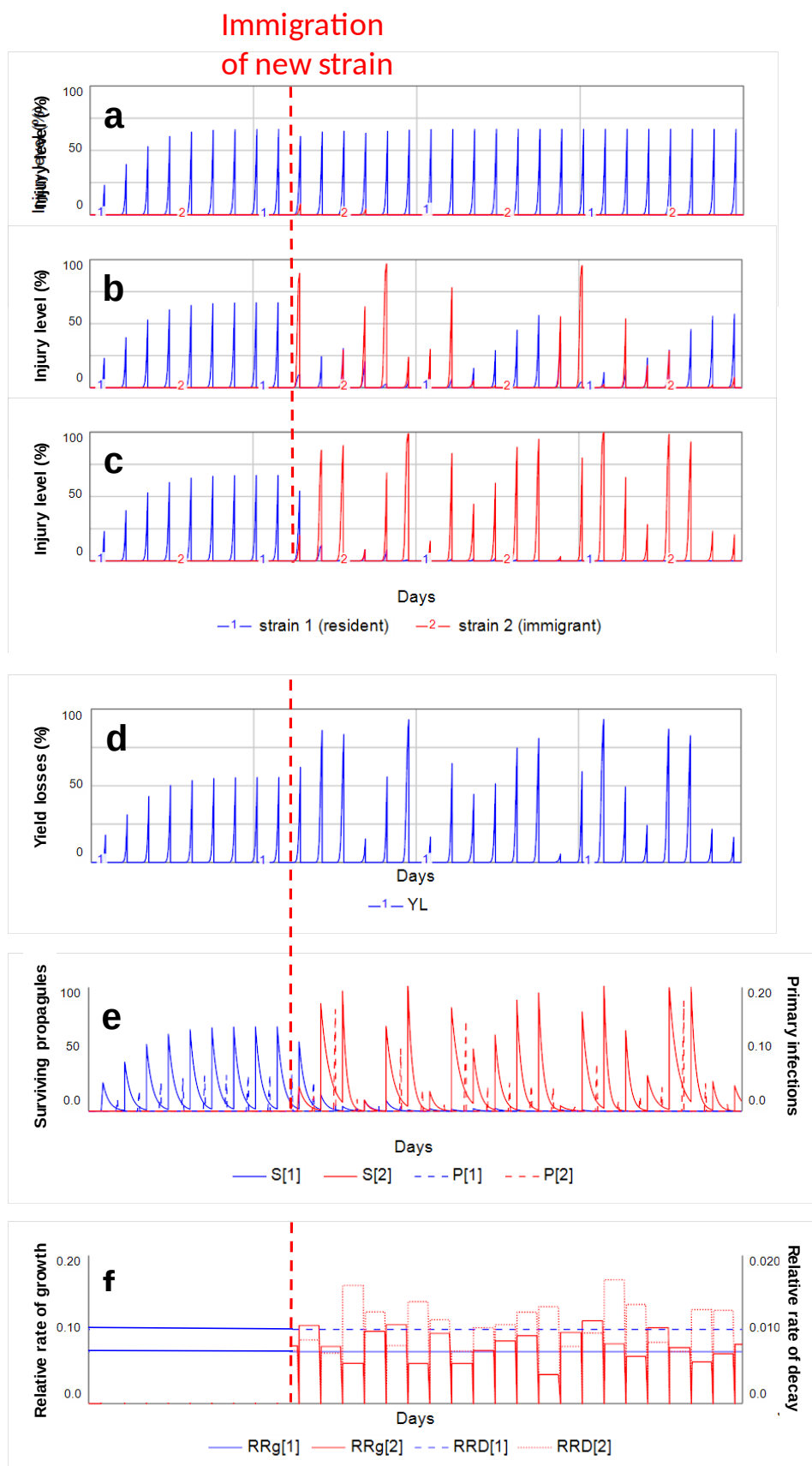
Acronym	Definition	Dimension	Unit	Value
State variables:				
<i>i</i>	Injury caused by disease on a crop stand	[-]	%	0 to 100
<i>S</i>	Number of surviving propagules	[-]	%	0 to 100
<i>YL</i>	Yield loss – yield reduction from a disease-free attainable yield	[-]	%	0 to 100
Rates:				
<i>RPI</i>	Rate of primary infections	[T <sup>-1</sup> ]	% Day <sup>-1</sup>	
<i>RconvIS</i>	Rate of conversion from injury (i) into surviving propagules (S)	[T <sup>-1</sup> ]	% Day <sup>-1</sup>	
<i>Rdecay</i>	Rate of decay of surviving propagules	[T <sup>-1</sup> ]	% Day <sup>-1</sup>	
<i>resetYL</i>	Rate of reset of yield loss at each cropping cycle	[T <sup>-1</sup> ]	% Day <sup>-1</sup>	
<i>RI</i>	Rate of injury increase	[T <sup>-1</sup> ]	% Day <sup>-1</sup>	
<i>RL</i>	Rate of yield loss	[T <sup>-1</sup> ]	% Day <sup>-1</sup>	
<i>RM</i>	Rate of immigration of the pathogen	[T <sup>-1</sup> ]	% Day <sup>-1</sup>	
<i>starter</i>	Rate of infection to initiate epidemics	[T <sup>-1</sup> ]	% Day <sup>-1</sup>	
Parameters:				
<i>CEP</i>	Duration of the crop establishment period	[T]	Day	120
<i>CGP</i>	Duration of the crop growth period	[T]	Day	120
<i>convSP</i>	Conversion of surviving propagules into a rate of primary infections	[T <sup>-1</sup> ]	Day <sup>-1</sup>	0.01
<i>imax</i>	Carrying capacity of injury – maximum level of injury	[-]	%	100
<i>RRD<sub>j</sub></i>	Relative rate of decay of surviving propagules for strain j	[T <sup>-1</sup> ]	Day <sup>-1</sup>	j = 1: 0.01 j = 2: Varying
<i>RRg<sub>j</sub></i>	Relative rate of epidemic (injury) increase for strain j	[T <sup>-1</sup> ]	Day <sup>-1</sup>	j = 1: 0.07 j = 2: Varying
<i>RRL</i>	Relative rate of yield loss	[T <sup>-1</sup> ]	Day <sup>-1</sup>	0.05
<i>Ya</i>	Attainable yield – uninjured yield level	[-]	%	100



**Fig. 1.** A framework for analysis of emerging epidemics: paths and processes. Three paths for emergence are considered (left, bold characters), involving different processes (in boxes). Paths may be combined, e.g., paths 1 and 3, involving both introduction and evolution, or 2 and 3, involving environmental change and evolution. See text for examples.



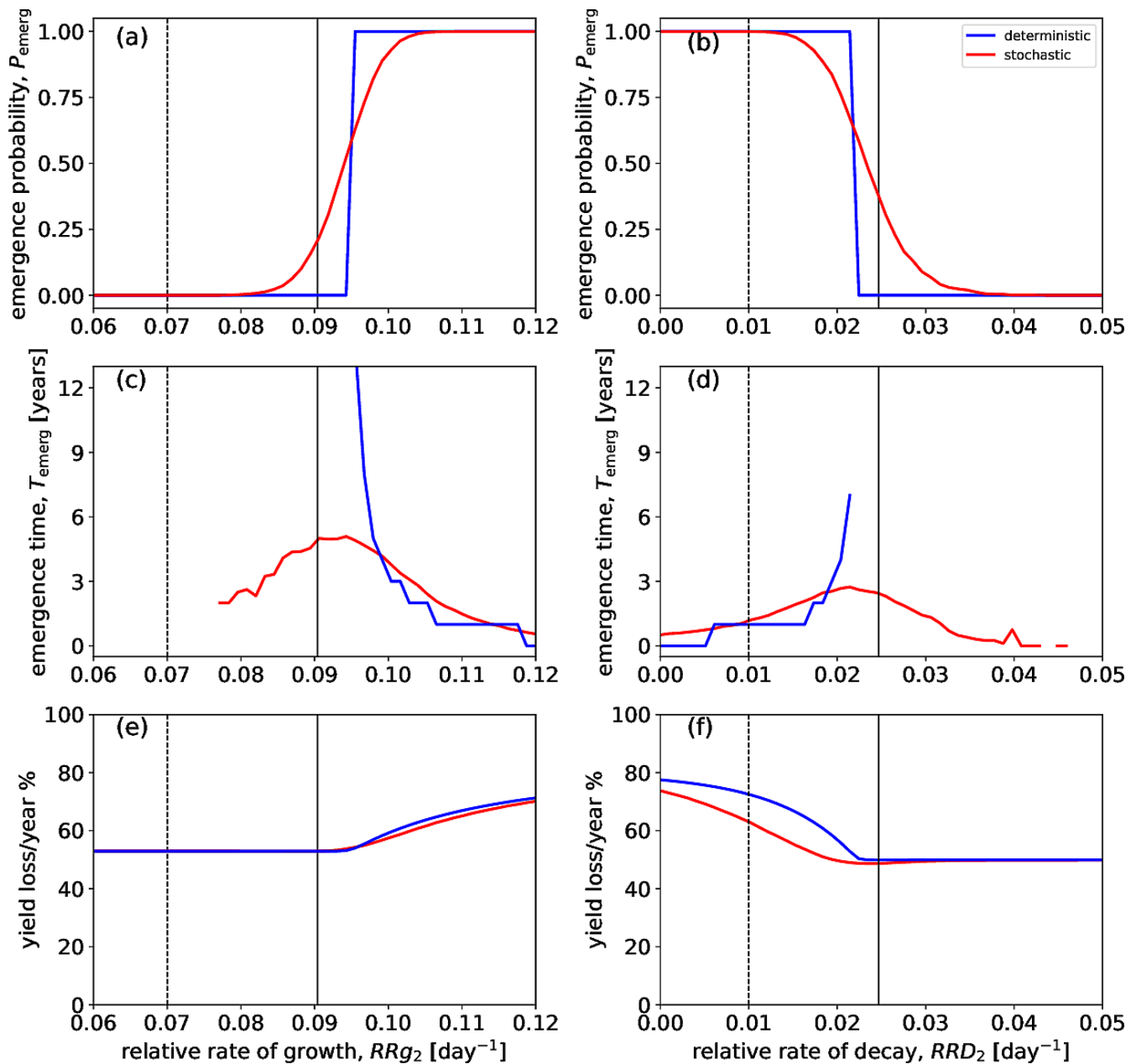
**Fig. 2.** Simplified flowchart of the process-based model used to analyse emerging diseases of crop plants. Variable acronyms are described in Table 1. The flowchart uses symbols introduced by J Forrester (Forrester, 1961): rectangles represent state variables; valves represent rates of change of state variables; circles represent parameters or computed variables. Stacked symbols (e.g., state variables) represent vectors of two pathogen strains.



**Fig. 3.** Examples of simulated dynamics of injury levels and related variables over 30 cropping cycles.

(a) to (c): simulated injury levels from different runs where  $RRg_2 \sim N(0.07, 0.03)$  and  $RRD_2 \sim N(0.01, 0.003)$ . (a): a simulation where the immigrant strain did not emerge; (b): a simulation where the immigrant and resident strains compete over cropping cycles; (c): a simulation where the immigrant strain replaces the resident strain; (d) to (f): simulated dynamics of other variables, associated with the dynamic of injury levels shown in (c).

Except for  $RRg_2$  and  $RRD_2$ , all parameters are set according to Table 1. Immigration of strain 2 takes place at the end of the crop establishment period of cropping cycle 10, while strain 1 is established at the beginning of the simulation.



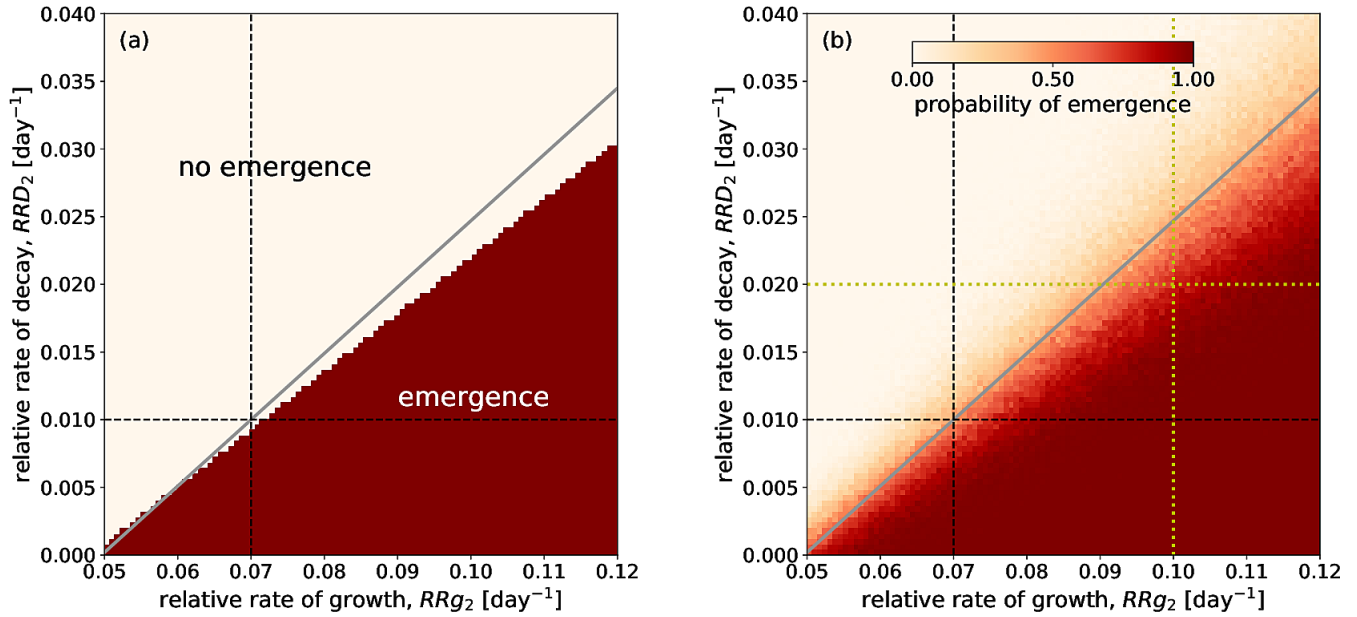
**Fig. 4.** Effects of pathogen reproduction and survival parameters on disease emergence and yield loss.

Emergence probabilities (a), (b), emergence times (c), (d) and yield losses (e), (f) are plotted versus the relative rate of growth,  $RRg_2$  (a), (c), (e), and the relative rate of decay,  $RRD_2$  (b), (d), (f), of the immigrant strain.

Blue curves correspond to the deterministic regime, in which both  $RRg_2$  and  $RRD_2$  have fixed, deterministic values, whereas red curves correspond to the stochastic regime. Red curves in panels (a), (c), (e) were computed with the values of  $RRD_2$  drawn as random numbers from the normal (Gaussian) distribution at the beginning of each cropping cycle with the mean  $0.02 \text{ day}^{-1}$  and the standard deviation  $0.007 \text{ day}^{-1}$ , while  $RRg_2$

assumed fixed, deterministic values. Similarly, red curves in panels (b), (d), (f) were computed with the values of  $RRg_2$  drawn as random numbers from the normal (Gaussian) distribution at the beginning of each cropping cycle with the mean  $0.1 \text{ day}^{-1}$  and the standard deviation  $0.035 \text{ day}^{-1}$ , while  $RRD_2$  assumed fixed, deterministic values.

Dashed vertical lines indicate the value of the relative rate of growth,  $RRg_1=0.07$ , of the resident strain in (a), (c), (e), and the value of the relative rate of decay,  $RRD_1=0.01$ , of the resident strain in (b), (d), (f). Solid vertical lines show the approximate analytical emergence thresholds, according to Eq. (6) in (a), (c), (e) and Eq. (7) in (b), (d), (f). Values of other parameters are given in Table 1.

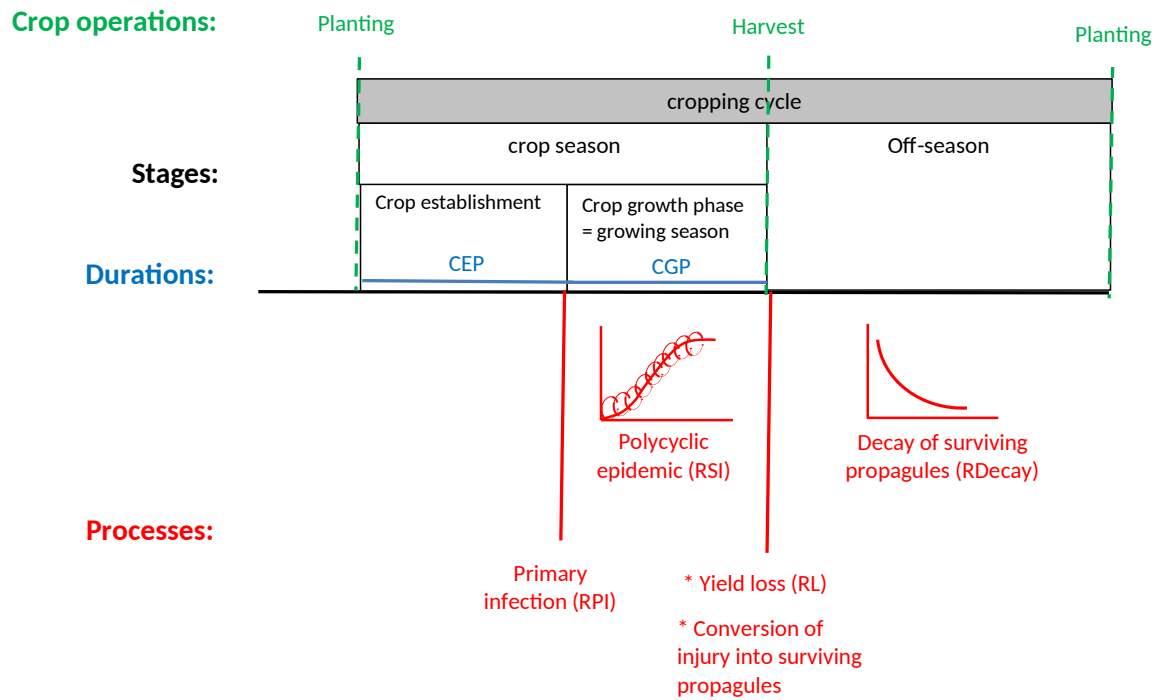


**Fig. 5.** Combined effects of pathogen reproduction and survival parameters on probability of emergence.

(a) deterministic regime; (b) stochastic regime, whereby the values of  $RRg_2$  and  $RRD_2$  were drawn from the normal (Gaussian) distribution at the beginning of each cropping cycle with the mean corresponding to the values on x- and y-axes and the standard deviation constituting a constant proportion, 0.35, of the corresponding mean values. Values of other parameters are given in Table 1.

In each panel, the grey diagonal line shows the analytical threshold according to Eq. (7). Dashed black lines indicate the fitness values of the resident strain: vertical  $RRg_1=0.07 \text{ day}^{-1}$ ; horizontal  $RRD_1=0.01 \text{ day}^{-1}$ . Dashed yellow lines mark the parameter regions explored in Fig. 4:  $RRD_2=0.02 \text{ day}^{-1}$  in Fig. 4a, c, e;  $RRg_2=0.1 \text{ day}^{-1}$  in Fig. 4b, d, f.





**Supplementary Figure 1.** Main stages and processes considered in the polyetic model at each

cropping cycle. CEP: duration of the Crop Establishment Phase; CGP: duration of Crop Growth phase.