Social copying drives a tipping point for nonlinear population collapse

Supplementary Information

Daniel Oro[1,2], Lluís Alsedà[3,4], Alan Hastings[2,5], Meritxell Genovart[1], and Josep Sardanyés[4]

[1]Theoretical and Computational Ecology Laboratory, Centre d'Estudis Avançats de Blanes (CEAB-CSIC), Cala Sant Francesc 14, 17300 Girona, Spain
[2]Department of Environmental Science and Policy, University of California, Davis, CA 95616, USA
[3]Departament de Matemàtiques, Edifici C, Universitat Autònoma de Barcelona, 08193 Bellaterra (Barcelona), Spain
[4]Centre de Recerca Matemàtica, Campus de Bellaterra, Edifici C, 08193 Bellaterra, Spain
[5]Santa Fe Institute, Santa Fe, NM 87501, USA

Contents

Supplementary Section 1

Ecological data: Audouin's gull study population

1.1 Study site and species

Long-term population monitoring of Audouin's gulls (*Ichthyaetus audouinii*) at Punta de la Banya (Ebro river Delta: $40°34'10.89''$N, $0°39'34.28''$E) started in 1981 when the patch was colonized and it has been continuously carried out until present during almost four decades (1981–2021). La Banya is a 2500ha sandy peninsula covered by halophilous vegetation and connected with the rest of the Ebro Delta by a 6km long narrow barren bar. The patch is at the mouth of the Ebro River and the continental shelf here is wide, which drives a high marine productivity and the overlap of high densities of both marine top predators and human fisheries [1, 2]. Audouin's gulls are long-lived social birds with a bet-hedging life history. They are philopatric but have evolved to cope with ephemeral habitats typical of Mediterranean marshes [3]. As a consequence, they have nomadic behaviour between breeding seasons: when patch conditions change and worsen then individuals are more prone to disperse mainly to other occupied sites [4], but in recent years and after the arrival of invasive carnivores, colonization rate of new patches has largely increased [5]. At La Banya, Audouin's gulls breed in sympatry with nine other species of the same ecological guild, including other gulls and terns. Dynamics and structure of this community is driven by competition governed by body size, with Audouin's gulls among the largest (i.e. dominant) species [6]. Demographic parameters of the species (by age and sex) have been estimated in a bunch of studies, including survival, recruitment curves, fertility and dispersal. Gulls breeding at La Banya show high adult survival (0.898 (SE : 0.01) mean adult survival probability), fast recruitment (birds start to breed when 3y old and most birds recruit before 5y old), and relative low fertility (average 0.471 fledglings per breeding female (SD : 0.287)) [4, 7–10]. Audouin's gulls are specialized predators to feed on small pelagics at night, but they have learnt to exploit fish discarded by trawlers, which provides up to 70% of their diet by biomass. Despite the noise caused by the presence of carnivores and some extreme climatic events, trawling discards explains 24% of the variance in fertility over the years (phase 1991–2017).

1.2 Fieldwork methods and environmental data

At La Banya, gulls (both Audouin's gulls and their main competitor yellow-legged gulls) build their nest in clumped groups (what is called a sub-colony) [11]. Censuses are performed by teams of 2–15 people depending on the size of the sub-colony. Those people are organized in parallel band strips of 2–3 meters of width and each person counts nests with eggs at the right part of her band, with the spatial limit imposed by the person counting at her right. People moved in parallel to avoid double nest counts and missing nests. Censuses are carried out during the late incubation phase before hatching to avoid biases due to individuals still incorporating to the reproductive season. Additional fieldwork details are explained elsewhere and census errors were estimated and considered small and constant over the years ($< 5\%$) [12]. Several biotic and abiotic drivers can influence population fluctuations at the study patch. However, previous studies show that local biotic drivers explain better these fluctuations than global oceanographic indexes, such as the North Atlantic Oscillation index NAO [9]. Among these biotic drivers, interference competition with the dominant yellow-legged gull and predation and disturbance by invasive carnivores (mainly foxes) are the main factors affecting all crucial demographic parameters, namely adult survival, bycatch, fertility and dispersal (both immigration and dispersal at spatial mesoscale). The main difference between these two drivers is that yellow-legged gulls are competitors with a long shared evolutionary history and long-term stability occurs when the two species occur in a specific patch. On the contrary, gulls have not developed evolutionary defenses to cope with terrestrial predators like carnivores, and this is why they select for breeding patches isolated and protected against the invasions of the predators. Population density of yellow-legged gulls and the number of carnivores present at La Banya have been estimated over the years (Figure S1.A,B, respectively), and gull carcasses and tracks in the sand allowed us to estimate yearly predation rates that varied with the individual predator and its foraging preferences [5, 11, 13]. Other biotic factor is food availability, and a proxy to assess its temporal variability is through the statistics of landings of trawlers in the harbors close to the study site, which are highly correlated with the amounts of fish discarded [2, 8]. To account for the strength of density-dependence, this proxy of the changes in food availability was transformed as food per capita by dividing by the sum of the densities of Audouin's and yellow-legged gulls, the two more abundant and dominant species in the community. This density-dependence index explains much of the variance in fertility (see above and Figure S1.C) and juvenile survival, whereas it did not correlate with changes in recruitment and adult survival [8, 9]. Food per capita decreased as population density approached the carrying capacity during the mid 90's and also because trawler catches per unit effort have decreased in recent decades due to overharvesting of fish stocks (Figure S1.D). Adult survival, which is the vital rate with largest elasticity for the population dynamics of the gulls, changes with bycatch mortality at longline fisheries and by carnivore predation [10, 14].

Previous studies have shown that bycatch is relatively constant over the years [15], whereas carnivore density may vary with breeding season, although values were always low (median number of adult carnivores since their first arrival equaled two with range between zero and five) [11]. Predation rate increased with the density of carnivores, but some noise for this association occurred due to individual carnivore preferences for gull predation (Figure S1.B). However, these predation rates did not significantly affect adult survival [4, 10], whereas they increased dispersal probabilities to other patches (either occupied or empty) [5, 11]. The number of colonized patches increased nonlinearly since 2006 (Figure S1.E), and metapopulation density followed parallel population dynamics with that at La Banya, except for the last years, when the slope at the former was slower than the slope to patch extinction at the later (Figure S1.F). In summary, we did not record a decrease of food availability in absolute and per capita values (i.e. accounting for density-dependence), nor a decrease of local survival by carnivore predation or an increase of competition with the dominant yellow-legged gulls. Thus, these variables cannot explain the decline of population density of Audouin's gulls to patch collapse at La Banya since 2006, which should respond to an increase of dispersal to other patches, previously recorded using marked individuals and their field monitoring along most of the whole western Mediterranean metapopulation [5, 10, 11].

1.3 Population trends

The geometric mean of the population growth rate expressed as $\ln\big(N_{t+1}/N_t\big)$, being N_t the size of the population at time t, of Audouin's gulls since colonization directly estimated from the field data in 1981 to 2017 was 0.086 (Figure S2). Annual gulls' mortality rate computed from long-term monitoring data (1988–2015) using capture-recapture modeling was estimated to be 0.11 year^{-1} [10]. Four well-defined periods can be distinguished when looking at the time series of breeding gulls, which are shown with vertical dashed lines in the time series of Figure S3. First, an *initial phase* with exponential population growth partly explained by high immigration rates from the outside [12]. Following this period, the population kept growing until 1997, when predators entered into the patch. The second period, labelled *onset of perturbation phase*, spans from 1998 to 2006. Figure S1.B displays the density of predators (blue dots) during the whole period of study, showing the presence of few predator individuals between 1997 and 2017, with a maximum of 5 individuals in 2010 and absence of carnivores from 2017 to 2021. The predation rate is also shown in Figure S1.B with red dots. Here, predation rate is the percentage of corpses predated by the carnivores with respect to the total number of corpses found each year (see [5, 11] for details). Despite some predation effect, the total number of corpses found dead by any cause during 1997–2017 was 703 with annual median 31 and annual range 14–87 (c.f. [5]). The population of gulls suffered a large increase around 2005–2006 due to an increase in food availability per capita (see also Figure S1.C,D). Since 2006, gull's population started a sustained and sharp decline until the patch held only 3% of total world population in 2017 (see main text), coinciding with the absence of predators in the patch (Figure S1.B). This period from 2006 to 2017 will be denoted as *collapse phase*. Finally, the period from 2017 to 2021, where the gulls' population started increasing again, coinciding with the absence of predators. In this manuscript we will focus on the dynamics between 1981 and 2017. The field data for the Audouin's gulls at Punta de la Banya during the period of study is shown in the table and in Figure S3 (see also Figure S1.A, green dots).

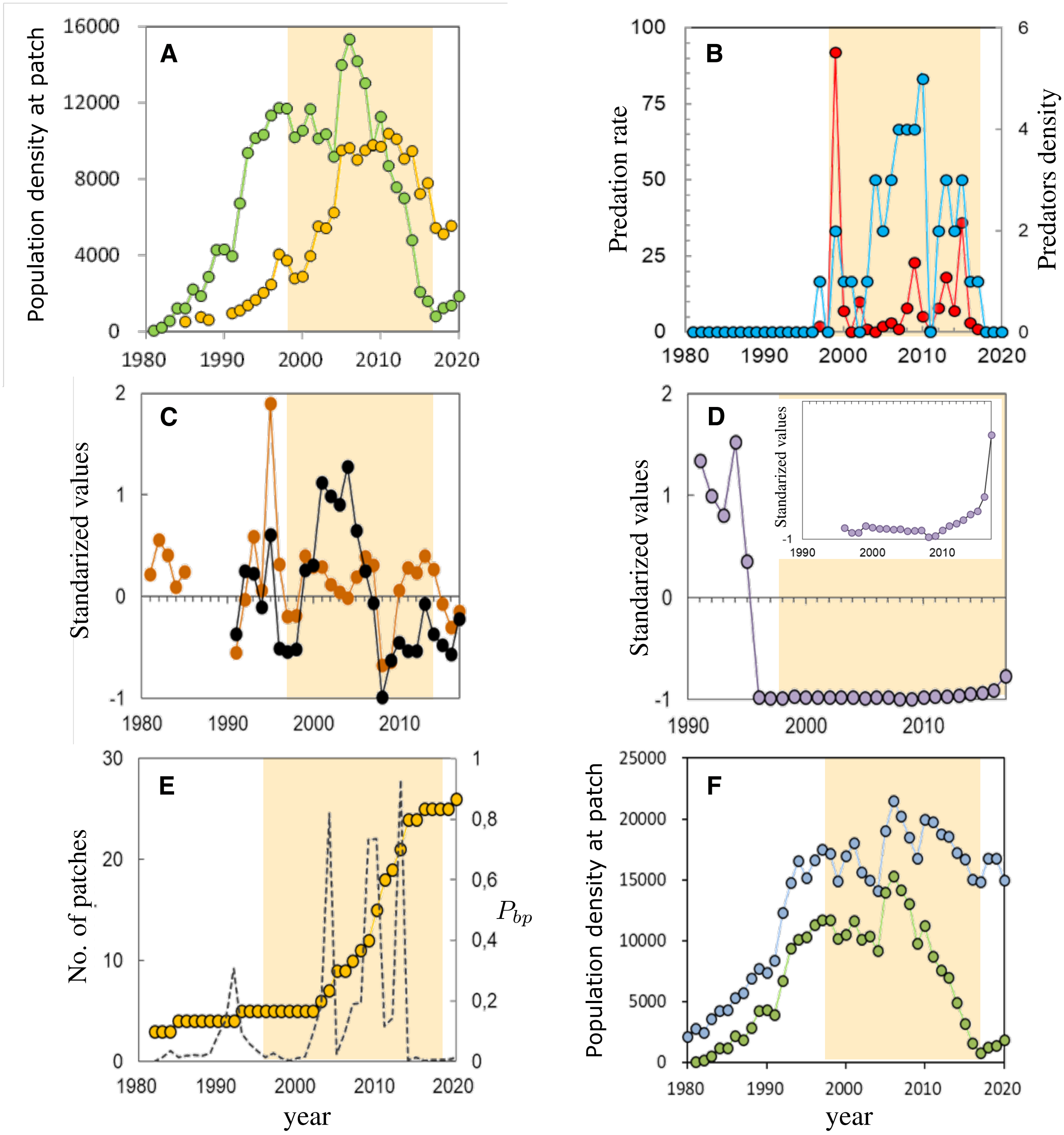


Figure S1. A: Population density of yellow-legged (yellow circles) and Audouin's gulls (green circles), at la Banya since 1981, when the later species colonized the patch.
B: Variability of predation rates by carnivores (as percentage of corpses found preyed at the patch relative to the total corpses found, red dots) and carnivore density (as number of adult carnivore present, blue dots).
C: Variability of trawling discards, as a proxy of food availability for gulls (maroon dots) and fertility (as mean number of chicks per breeding female, black dots) at La Banya.
D: Variability of food availability per capita during 1991–2017 and for the phase of population of Audouin's gull attaining the carrying capacity (inner panel).
E: Accumulated number of breeding patches occupied in the western Mediterranean and southern Portugal since 1981 (circles) and Bayesian probability of detecting a breaking point for this time series (dashed line).
F: Population density of Audouin's gulls at La Banya (green circles) from colonization in 1981 to 2020 and metapopulation density for this species including all patches in the western Mediterranean and southern Portugal (blue circles).

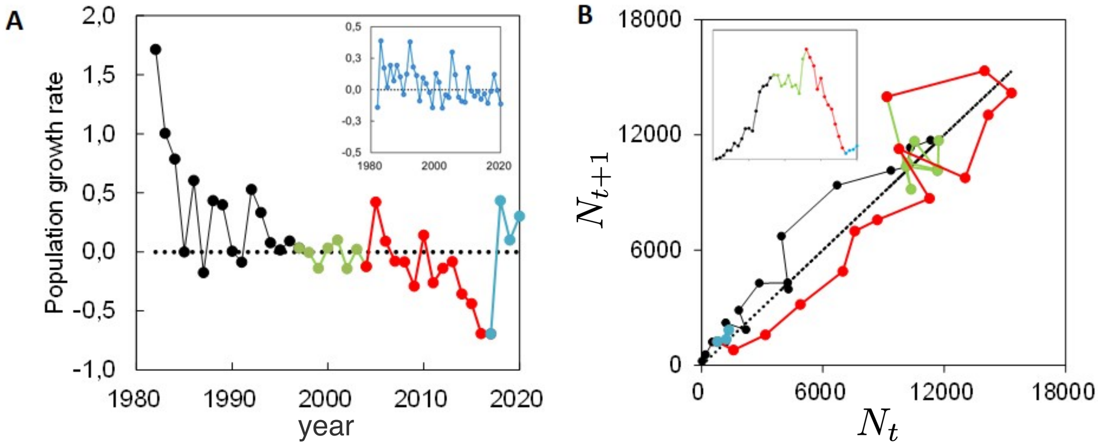


Figure S2. **A:** Population growth rate since colonization of La Banya by Audouin's gulls in 1981; the dashed line shows no population growth. The inner panel shows population growth rate for the whole metapopulation (90% of total world population); black, green and red colors show the phases of exponential initial growth, dynamic stability, and nonlinear decline respectively.
B: Ricker function of population density N at time t versus time $t+1$, with dashed line showing stability; colours as in panel (A); the inner panel shows how population density of Audouin's gulls varied at La Banya since colonization to 2020.

year	pop.	year	pop.	year	pop.
1981	36	**1995**	10327	**2009**	9762
1982	200	**1996**	11328	**2010**	11271
1983	546	**1997**	11725	**2011**	8688
1984	1200	**1998**	11691	**2012**	7571
1985	1200	**1999**	10189	**2013**	6983
1986	2200	**2000**	10537	**2014**	4778
1987	1850	**2001**	11666	**2015**	2067
1988	2861	**2002**	10122	**2016**	1586
1989	4266	**2003**	10355	**2017**	793
1990	4300	**2004**	9168	**2018**	1225
1991	3950	**2005**	13988	**2019**	1355
1992	6174	**2006**	15329	**2020**	1837
1993	9373	**2007**	14177	**2021**	2319
1994	10143	**2008**	13031		

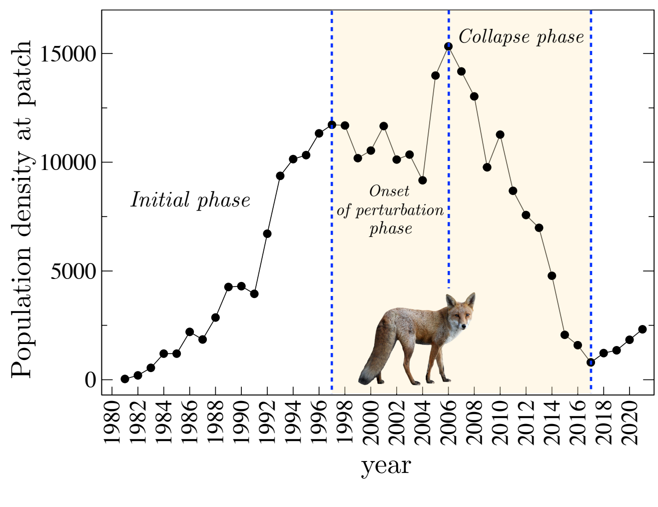


Figure S3. The Audouin's population data at La Banya from 1981 to 2021. The period 1981–1997, labelled as *initial phase*, was characterised by a logistic growth due to the absence of predators and the fact that the population did not approach the expected equilibrium (see Section S3.2.2 below). Predators (foxes) colonized the patch in 1997, causing a qualitative change in the dynamics and a decreasing tendency in the population until 2004, with a large fluctuation in 2005–2006. The period 2006–2017, labelled as *collapse phase*, was characterised by a fast population decline due to dispersal until 2017, when predators were not found at all in the patch. Notice that after the absence of predators from 2017 onwards, the population of birds started increasing again (2018–2021).

Supplementary Section 2

Mathematical modeling with dispersal by social copying

In this section we introduce the mathematical model used to investigate the population dynamics of Audouin's gulls in the patch of study. The model describes the population dynamics of the birds as a single-patch system considering immigration and dispersal of individuals. Other modelled ecological processes are birds' intra-specific competition for resources and density-independent death. As we thoroughly explain below, the model incorporates dispersal dynamics considered in the last two terms at the right hand side of Equation (2.1). Let us denote the birds' population size at time t by $x(t)$. Then, the model reads:

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}t}\,x(t) &= (\vartheta+\omega)x(t)\left(1-\frac{x(t)}{K}\right)-\varepsilon x(t)-\Big[\rho x(t)+\lambda D\big(x(t)\big)\Big]\\ &= \underset{\text{Immigration, growth and death}}{\boxed{\varphi x(t)}} - \underset{\text{Nonlinear competition term}}{\boxed{\beta x(t)^2}} - \underset{\text{Dispersal by social copying}}{\boxed{\lambda D(x(t))}}\end{aligned} \tag{2.1}$$

where the parameters are described in Table 1 in the next page.

Equation (2.1) considers an initial exponential increase of the population proportional to parameters $\vartheta+\omega$, including both the reproduction of birds (ϑ) and the immigration rate (ω) of new individuals from other patches of the metapopulation (not explicitly considered) to the patch of study. Notice that, for simplicity, we have lumped these two parameters by setting $\gamma=\vartheta+\omega$. The population growth is constrained by a logistic function with carrying capacity K, introducing intra-specific competition for resources. The competition term can also be expressed as $\beta x(t)^2$, with $\beta=\gamma/K$. The death rate is fixed to $\varepsilon=0.11$ corresponding to the annual mortality rate estimated from long-term monitoring (1988–2015) using capture-recapture modeling [10]. Finally, two terms related to dispersal are included. Exponential (positive density-dependent) dispersal proportional to constant ρ and the function $D(x(t))$ that will be used to introduce dispersal by social copying. This dispersal term mimicking social copying will generically consider a negative density-dependent departure of the birds from the patch. That is, the less number of birds at the patch the higher their departure (see Section S4).

The model in compact form is shown framed with different colours: the blue box displays all of the processes related to population growth, including reproduction, immigration, death, and exponential dispersal of individuals now with $\varphi=\gamma-\varepsilon-\rho$. The green box corresponds to the intra-specific competition, while the dispersal term including social copying is represented within the red box. The model will be used to investigate the field data, focusing on three phases. We call the period 1981–1997, which corresponds to the establishment of the local population before the arrival of predators (i.e., before the perturbation), the *initial phase*. It is characterised by a logistic growth due to the absence of predators and the fact that the population did not exhaust the food carrying capacity. Here, we will not consider dispersal triggered by the perturbation since predators were not present in the patch during this period (i.e., $\rho=\lambda=0$). The second period, labeled *onset of perturbation phase*, will incorporate the hypothesized social copying dynamics triggered by the arrival of predators in the patch, and will last till the beginning of the collapse phase starting in 2006. This phase is characterised by a change in the growing tendency of the population in 1998 and a sustained decrease until 2004, with a very large increase in the years 2005–2006. Since the available data for this period is scarce, we will use the parameters of the Elliot sigmoid function estimated from 2006 to 2017 to fit this second phase leaving as free parameters ρ and λ (Section S4.3). We will investigate the dynamics from 2006–2017 (*collapse*

Parameter	Units	Range or value	Ecological meaning or description
ϑ	year^{-1}	$[0, +\infty)$	Intrinsic reproduction rate
ω	year^{-1}	$[0, +\infty)$	Rate of entry of individuals from other patches
K	birds	$[1, +\infty)$	Carrying capacity
ε	year^{-1}	0.11	Death rate estimated from field data [10]
ρ	year^{-1}	$\mathbb{R}^+$	Linear (exponential) dispersal rate
$\gamma = \vartheta + \omega$	year^{-1}	$[0, +\infty)$	Population growth rate due to reproduction and immigration
$\alpha = \gamma - \varepsilon$	year^{-1}	$(-\infty, \gamma]$	Neat population growth rate without *linear* dispersal
$x(0)$	birds	$[0, K]$	Initial condition of Equation (2.1)
$\varphi = \alpha - \rho$	year^{-1}	$(-\infty, \alpha]$	Population growth rate including linear dispersal
$\beta = \frac{\gamma}{K}$	(birds $\times$ year)$^{-1}$	$[0, +\infty)$	Intrinsic growth rate over the carrying capacity
λ	year^{-1}	$\mathbb{R}^+$	Dispersal rate by social copying

Table 1. Model parameters for the general model used to investigate the local dynamics of Audouin gulls at la Punta de la Banya from 1981 to 2017.

phase), when predators were still present at the patch and the gulls' population experienced the collapse, with Equation (2.1). That is, considering dispersal terms. We want to emphasize that we will mainly focus on the collapse phase from 2006 to 2017, since the fitting for the period 1998 to 2004 contains very few data. The analytic and qualitative study of the model dynamics for these three phases and the corresponding fitting of parameters will be done in the next sections.

Supplementary Section 3

Dynamics before the perturbation

3.1 Model dynamics

In this section we study analytically and qualitatively the model given by Equation (2.1) in the initial phase, ranging from the establishment of the population at the patch of study in 1981 until the arrival of predators in 1997. The model given by Equation (2.1) to study the initial phase will not include dispersal by social copying since we hypothesize that this behavioural dispersal is triggered by the presence of predators ($\lambda = 0$). Moreover, we will also assume no linear dispersal from the study patch to other patches of the metapopulation ($\rho = 0$) since, as we discussed in Section 1.2 above, the initial phase was dominated by high immigration rates from outside [12]. Under these considerations, we get

$$\frac{dx(t)}{dt} = \alpha x(t) - \beta x(t)^2. \tag{3.1}$$

Equation (3.1) is a particular case of a Ricatti Equation with constant coefficients. Its closed analytical solution is obtained in the next lemma by integrating the model as an equation of separable variables.

Lemma 3.1 (A Ricatti Equation with constant coefficients)**.** *According to Table 1 above we know that β must be non-negative. Then, the solution $x(t)$ of Model* (3.1) *is the following:*

If $\alpha \neq 0$,

$$x(t) = \frac{\alpha x(0)\exp(\alpha t)}{\alpha + \beta x(0)(\exp(\alpha t) - 1)} = \frac{\alpha x(0)}{\alpha \exp(-\alpha t) + \beta x(0)(1 - \exp(-\alpha t))};$$

and if $\alpha = 0$,

$$x(t) = \frac{x(0)}{x(0)\beta t + 1}.$$

The dynamics of Model (3.1) is well-known. However, for the sake of completeness we here analyse its qualitative dynamics for the case $\alpha, \beta > 0$. This is indeed not a restrictive assumption since as we will see below the observed data is only compatible with the positivity of parameters α and β.

The proof of the next lemma is a simple exercise (see Figure S4).

Lemma 3.2. *Assume that $\alpha, \beta > 0$. Then the function $f(x) = x(\alpha - \beta x)$ verifies: $f(0) = f\left(\frac{\alpha}{\beta}\right) = 0$, $f(K) = -K\varepsilon \leq 0$, and has a unique critical point at $x = \frac{\alpha}{2\beta}$. Hence, f is unimodal with $f\big|_{\left(0, \frac{\alpha}{2\beta}\right]}$ strictly increasing and positive, $f\big|_{\left[\frac{\alpha}{2\beta}, K\right]}$ strictly decreasing, $f\big|_{\left[\frac{\alpha}{2\beta}, \frac{\alpha}{\beta}\right)}$ positive and $f\big|_{\left(\frac{\alpha}{\beta}, K\right]}$ negative. Additionally, $0 < \frac{\alpha}{2\beta} < \frac{\alpha}{\beta} \leq K$, and $\frac{\alpha}{\beta} = K$ if and only if $\varepsilon = 0$.*

A consequence of Lemma 3.2 is that model (3.1) has two stationary solutions computed from $\dot{x} = 0$. They are $x(t) = 0$ and $x(t) = \frac{\alpha}{\beta}$ when $t \to +\infty$. Equilibrium 0 (labeled $x_0^\star$) involves, whenever stable, no population at the patch, while equilibrium $\frac{\alpha}{\beta}$ (labeled $x_1^\star$) will involve, provided is stable, the persistence of the population. Generically, the (local) stability of a given equilibrium solution $x(t \to +\infty) = x^\star$ of a one-variable differential equation $\frac{dx(t)}{dt} = f(x(t))$ can be computed from the sign of $\frac{df(x)}{dx}\big|_{x=x^\star}$. More precisely, the equilibrium is a local attractor when $\frac{df(x)}{dx}\big|_{x=x^\star} < 0$ or unstable when $\frac{df(x)}{dx}\big|_{x=x^\star} > 0$. From the previous expressions we obtain

$$\left.\frac{df(x)}{dx}\right|_{x=x^\star} = \alpha - 2\beta x^\star = \begin{cases} \alpha & \text{if } x^\star = x_0^\star = 0, \text{ and} \\ -\alpha & \text{if } x^\star = x_1^\star = \frac{\alpha}{\beta}. \end{cases}$$

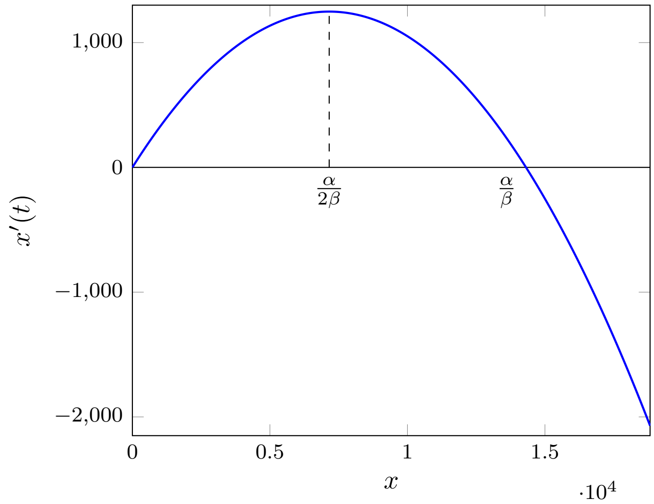


Figure S4. The vector field (3.1) for $\alpha = 0.3489494104672237$, $\beta = 2.43826356697 \times 10^{-5}$ and $K = 18822.8$.

Hence, for $\alpha > 0$ (intrinsic growth rate of the population larger than decline rate) $x_0^\star$ is unstable and $x_1^\star$ stable. We notice that for $\alpha = 0$ there exists a transcritical bifurcation involving a collision and an exchange of stability between the two equilibria. As expected, for $\alpha < 0$ (decline rate larger than population growth rate) the stable equilibrium corresponds to $x_0^\star = 0$. This behaviour is illustrated in Figure S5A as a function of α. As we will see in Section 3.2, the case of interest is given by $\alpha > 0$, value obtained for the field data.

Finally, an illustrative way of visualizing the stability of a dynamical system with one variable is to compute the so-called potential function, given by:

$$U(x) = -\int f(x)\, dx, \qquad \text{with} \quad U(x) = x^2 \left(\frac{\beta x}{3} - \frac{\alpha}{2} \right) \text{ for Equation (3.1).} \tag{3.2}$$

Figure S5B displays three potential functions computed from Equation (3.2) for different values of $\alpha > 0$. Specifically, the field data for the initial phase reveals that the equilibrium of the population was not achieved when predators colonised the patch, thus being in a transient state (see Figure 3B in the main manuscript and Section S3.2).

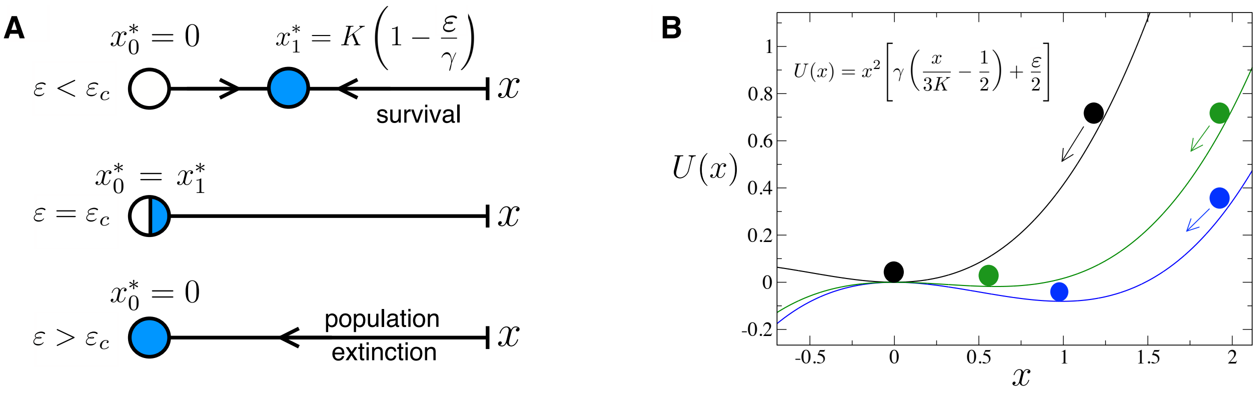


Figure S5. A: Schematic diagram of the dynamics of Equation (3.1) in the phase space as a function of model parameters. Here three possible scenarios are found: $\alpha < 0$ with population vanishment (x_0^* stable and x_1^* unstable); $\alpha = 0$, bifurcation value at which the transcritical bifurcation occurs; and $\alpha > 0$, with persistence of the population (x_1^* stable and x_0^* unstable). Stable and unstable points are indicated with blue and white marbles, respectively.
B: Potential function given by Equation (3.2) computed with $\alpha = 0.4995$ (blue); $\alpha = 0.3$ (green); $\alpha = -0.495$ (black). Here we use $K = 1$.

3.2 Model fitting and parameters estimation: Initial phase 1981–1997

3.2.1 On the positivity of structural parameters: analytical proof

For the study and fitting of Model (3.1) in the initial phase (Figure S3) we need to introduce some appropriate notation. The observed population of Audouin's gulls at the years 1981 to 1997 will be denoted by

$$\eta(\ell, \ell = 0:16) = \mathsf{Audouin's\_Gulls\_Observed\_Population\_at\_year}(1981 + \ell, \ell = 0:16) =$$
$$\big[\overset{[0]}{36}, \overset{[1]}{200}, \overset{[2]}{546}, \overset{[3]}{1200}, \overset{[4]}{1200}, \overset{[5]}{2200}, \overset{[6]}{1850}, \overset{[7]}{2861}, \overset{[8]}{4266}, \overset{[9]}{4300}, \overset{[10]}{3950}, \overset{[11]}{6714}, \overset{[12]}{9373}, \overset{[13]}{10143}, \overset{[14]}{10327}, \overset{[15]}{11328}, \overset{[16]}{11725}\big].$$

The solution of the above Model (3.1) with $\beta \geq 0$ and initial condition $\kappa \in \mathbb{R}^+$ will be denoted by $x(t) = x_{\kappa,\alpha,\beta}(t)$. Observe that $\kappa = x_{\kappa,\alpha,\beta}(0)$ must be considered a free parameter as well.

Now we define the parameter space

$$\mathscr{F} := \mathbb{R}^+ \times \big\{(\beta K - \varepsilon, \beta) : \beta \in \mathbb{R}^+\big\}$$

(recall that $\alpha = \gamma - \varepsilon = \beta K - \varepsilon \in (-\varepsilon, \infty)$), and a map

$$\begin{array}{rccc} \mathsf{L}\colon & \mathscr{F} & \longrightarrow & \mathbb{R}^+ \\ & \big(\kappa, (\alpha, \beta)\big) & \longmapsto & \sqrt{\sum_{\ell=0}^{16} \big(x_{\kappa,\alpha,\beta}(\ell) - \eta(\ell)\big)^2}. \end{array}$$

The map L measures the agreement between the solution of Model (3.1) with initial condition κ and parameters α and β, and the observed data $\eta(\ell, \ell = 0:16)$, through the Euclidean norm:

$$\sqrt{\sum_{\ell=0}^{16} \big(x_{\kappa,\alpha,\beta}(\ell) - \eta(\ell)\big)^2}.$$

The Euclidean norm of the errors will be systematically used throughout our analyses and will be labeled as quadratic error or least-squares due to its similarity with regression theory.

Observe that the map $\big(\kappa, (\alpha, \beta)\big) \longmapsto \sqrt{\sum_{\ell=0}^{16} \big(x_{\kappa,\alpha,\beta}(\ell) - \eta(\ell)\big)^2}$ can be decomposed in two steps:

$$\big(\kappa, (\alpha, \beta)\big) \longmapsto x_{\kappa,\alpha,\beta}(\ell, \ell = 0:16) \longmapsto \sqrt{\sum_{\ell=0}^{16} \big(x_{\kappa,\alpha,\beta}(\ell) - \eta(\ell)\big)^2}.$$

The first step is computed with the help of Lemma 3.1, taking into account whether $\alpha < 0$; $\alpha = 0$ or $\alpha > 0$.

Of course, if the dynamics of the Audouin's gulls population size during the years 1981 to 1997 is governed by some instance of Model (3.1) with parameters $x(0) = \kappa^*$, $\alpha = \alpha^*$ and $\beta = \beta^*$, then the value of $\mathsf{L}\big(\kappa^*, (\alpha^*, \beta^*)\big)$ must be small and likely it must correspond to

$$\begin{gathered} \min \mathsf{L}\big(\kappa, (\alpha, \beta)\big) \\ \text{subject to } \big(\kappa, (\alpha, \beta)\big) \in \mathscr{F}, \\ x(0) = \kappa \geq 0, \\ \text{and } x(t) \geq 0 \text{ for } t \in [0, 16]. \end{gathered} \tag{3.3}$$

The solution of this problem is called the *fitting of the model* and identifies a valid analytical model for the dynamics of the Audouin's gulls at the years 1981 to 1997 (of course provided that the value $\min \mathsf{L}\big(\kappa, (\alpha, \beta)\big)$ is small). Observe that the set

$$\Big\{\mathsf{L}\big(\kappa, (\alpha, \beta)\big) : \big(\kappa, (\alpha, \beta)\big) \in \mathscr{F}; x_{\kappa,\alpha,\beta}(0) = \kappa \text{ and } x_{\kappa,\alpha,\beta}(t) \geq 0 \text{ for } t \in [0, 16]\Big\} \subset \mathbb{R}^+$$

has 0 as a lower bound. Hence, it has a minimum element, and Problem (3.3) has at least one solution.

As mentioned above, we will use the quadratic error as a measure to check how good the fit of the population values obtained with the dynamic model is with respect to the real data. Hence, this will be the objective function to be minimized in the search for the best fitting parameters of the model. This quadratic error is a number between 0 (simulated data reproduce exactly real data) and $+\infty$. In addition, we will provide for the relevant cases a measure of correlation between the simulated and the real data, by means of the *coefficient of determination*, R^2. Here, the case $R^2 = 1$ will involve a complete agreement between simulated and real data (equivalent to least-squares equal zero) while $R^2 \sim 0$ will involve no correlation between these data. Generically, R^2 can be defined as:

$$R^2 = 1 - \frac{\text{Residual sum of squares}}{\text{Total sum of squares}} = 1 - \frac{\sum_{\ell=0}^{n-1} \big(x(\ell) - \eta(\ell)\big)^2}{\sum_{\ell=0}^{n-1} \Big(\eta(\ell) - \overline{\eta}\Big)^2},$$

where n denotes the number of data points, $\eta(\ell)$ denotes the field data, and

$$\overline{\eta} := \frac{\sum_{\ell=0}^{n-1} \eta(\ell)}{n}$$

is the field data average. For instance, for the fitting of the initial phase performed in the next section, the determination coefficient reads

$$R^2 = 1 - \frac{\Big(\mathsf{L}\big(\kappa, (\alpha, \beta)\big)\Big)^2}{\sum_{\ell=0}^{16} \Big(\eta(\ell) - \overline{\eta}\Big)^2}$$

Observe that the *Total sum of squares* $= \sum_{\ell=0}^{n-1} \Big(\eta(\ell) - \overline{\eta}\Big)^2$ is indeed n times the variance of the observed data. We will proceed in a similar manner for most of the fittings developed in the following sections. We want to emphasize that R^2 is here given as an orienting measure of the quality of the fit in terms of correlation between the simulated and the observed data. Such statistics must be applied for linear regression models, and does not provide statistical information for nonlinear problems as the one we are addressing in this study [16]. Hence, this measure is purely orienting.

Next we consider a reduced (and better) parameter space

$$\mathscr{O} := \mathbb{R}^+ \times \Big\{ \big(\alpha, \tfrac{\alpha+\varepsilon}{K}\big) : \alpha \in (0, \infty) \Big\} \subset \mathscr{F}$$

(here we use again that $\beta = \frac{\gamma}{K} = \frac{\alpha+\varepsilon}{K}$), and the associated reduced optimization problem becomes:

$$\begin{aligned} &\min \mathsf{L}\big(\kappa, (\alpha, \beta)\big) \\ &\text{subject to } \big(\kappa, (\alpha, \beta)\big) \in \mathscr{O}, \\ &x(0) = \kappa \geq 0, \\ &\text{and } x(t) \geq 0 \text{ for } t \in [0, 16]. \end{aligned} \tag{3.4}$$

The next lemma reduces the search space to find the optimum fit of the model and, since $\alpha > 0$ whenever $\big(\kappa, (\alpha, \beta)\big) \in \mathscr{O}$, Lemma 3.1 tells us that in the first step of the computation of the map L we have

$$x_{\kappa,\alpha,\beta}(t) = \frac{\alpha\kappa}{\alpha \exp(-\alpha t) + \beta\kappa(1 - \exp(-\alpha t))}.$$

Lemma 3.3. *The solutions of Problem* (3.3) *and Problem* (3.4) *coincide.*

Proof. We have to see that for every $\big(\kappa, (\alpha, \beta)\big) \in \mathscr{F} \setminus \mathscr{O}$ there exist $\big(\widetilde{\kappa}, (\widetilde{\alpha}, \widetilde{\beta})\big) \in \mathscr{O}$ such that

$$\mathsf{L}\big(\widetilde{\kappa}, (\widetilde{\alpha}, \widetilde{\beta})\big) < \mathsf{L}\big(\kappa, (\alpha, \beta)\big).$$

The following is obtained by direct computation: $\big(200, \big(0.3, 10^{-5}\big)\big) \in \mathscr{O}$, and

$$\mathsf{L}\big(200, \big(0.3, 10^{-5}\big)\big) = 6317.69 \cdots.$$

On the other hand, for every $\big(\kappa, (\alpha, \beta)\big) \in \mathscr{F} \setminus \mathscr{O}$ we have $\alpha \leq 0$ (in fact $\alpha \in [-\varepsilon, 0]$). Since β is non-negative, this implies $x'(t) = x(t)\big(\alpha - \beta x(t)\big) \leq 0$ because $x(t) \geq 0$ for every t. Then, for every $t \geq 0$, by the Mean Value Theorem, there exists a $\xi \in (t, t+1)$ such that $x(t+1) = x(t) + x'(\xi) \leq x(t)$. Consequently,

$$\begin{aligned} x_{\kappa,\alpha,\beta}(0) \geq x_{\kappa,\alpha,\beta}(1) \geq \cdots \geq x_{\kappa,\alpha,\beta}(6) \geq x_{\kappa,\alpha,\beta}(7) \geq x_{\kappa,\alpha,\beta}(8) \geq \\ x_{\kappa,\alpha,\beta}(9) \geq x_{\kappa,\alpha,\beta}(10) \geq x_{\kappa,\alpha,\beta}(11) \geq \cdots \geq x_{\kappa,\alpha,\beta}(16). \end{aligned}$$

Assume first that $3950 = \eta(10) \le x_{\kappa,\alpha,\beta}(7)$. Then

$$\eta(\ell) < \eta(10) \le x_{\kappa,\alpha,\beta}(7) \le x_{\kappa,\alpha,\beta}(\ell) \quad \text{for } \ell = 0, 1, \dots, 7.$$

Then,

$$\mathsf{L}\big(\kappa, (\alpha, \beta)\big) \ge \sqrt{\sum_{\ell=0}^{7} \big(x_{\kappa,\alpha,\beta}(\ell) - \eta(\ell)\big)^2} \ge \sqrt{\sum_{\ell=0}^{7} \big(\eta(10) - \eta(\ell)\big)^2} = 8046.89\cdots > \mathsf{L}\big(200, (0.3, 10^{-5})\big).$$

This ends the proof of the lemma in this case.

Now assume that $3950 = \eta(10) \ge x_{\kappa,\alpha,\beta}(7)$. In this case, $\eta(\ell) \ge \eta(10) \ge x_{\kappa,\alpha,\beta}(7) \ge x_{\kappa,\alpha,\beta}(\ell)$ for $\ell = 8, 9, \dots, 16$. Hence,

$$\mathsf{L}\big(\kappa, \alpha, \beta\big) \ge \sqrt{\sum_{\ell=8}^{16} \big(\eta(\ell) - x_{\kappa,\alpha,\beta}(\ell)\big)^2} \ge \sqrt{\sum_{\ell=8}^{16} \big(\eta(\ell) - \eta(10)\big)^2} = 15204.468\cdots > \mathsf{L}\big(200, (0.3, 10^{-5})\big).$$

□

3.2.2 Estimation of the structural population parameters

Here, we estimate the structural population parameters better explaining the dynamics of the initial phase. These include the initial condition, κ, and parameters α and β (see Table 1), taking the value of $\varepsilon = 0.11$ estimated from the field data [10]. In order to obtain the structural parameters of the population we need to focus on the local dynamics before the perturbation and not considering external perturbations. That is, focusing on the period 1981–1997. From the discussion in Section 3.1 (see Lemma 3.3) we have to solve Problem (3.4). To do so, we have used a standard *trust region method* and also the Levenberg-Marquardt algorithm to solve the trust region sub-problem (see the GNU Scientific Library (GSL) *Nonlinear Least-Squares Fitting* documentation). As it can be guessed from the last reference we have used a GSL standard library function for this computation, with numerical approximation of derivatives of the objective function. The obtained parameters and data corresponding to the optimum obtained are given below:

Parameters' values at the optimum

Parameter	Value
$\kappa = x_{\kappa,\alpha,\beta}(0)$	288.04096 ± 117.9663
α	$0.3489494104672237 \pm 0.04958259$
β	$0.0000243826356697 \pm 0.00000598145$
ε	0.11 (estimated from data [10])
$\gamma = \alpha + \varepsilon$	$0.45894940\cdots$
$K = \frac{\gamma}{\beta}$	$18822.79734\cdots$
Quadratic Error $= \mathsf{L}\big(\kappa, (\alpha, \beta)\big)$	$2593.053\cdots$
Coefficient of determination R^2	$0.97620\cdots$

Year	Population Data: Observed	Population Data: Predicted
1981	36	288.040962
1982	200	404.917270
1983	546	567.299150
1984	1200	791.095769
1985	1200	1096.137854
1986	2200	1505.703287
1987	1850	2044.623822
1988	2861	2735.234088
1989	4266	3590.827296
1990	4300	4607.530655
1991	3950	5757.502051
1992	6714	6987.807453
1993	9373	8228.125064
1994	10143	9405.848001
1995	10327	10462.226529
1996	11328	11362.442028
1997	11725	12096.688846

In the left part of Figure S6 we display the dynamics of the local population for the estimated parameter values shown above (red line). The predicted equilibrium, computed from

$$x_1^\star = \frac{\alpha}{\beta} = 14311.390089$$

is shown with a horizontal brown line, while the carrying capacity is shown with a horizontal dashed line. The model predictions suggest that the population have not reached the steady state on the onset of the perturbation, and that the large population increase suffered in 2005–2006 did not surpass the carrying capacity.

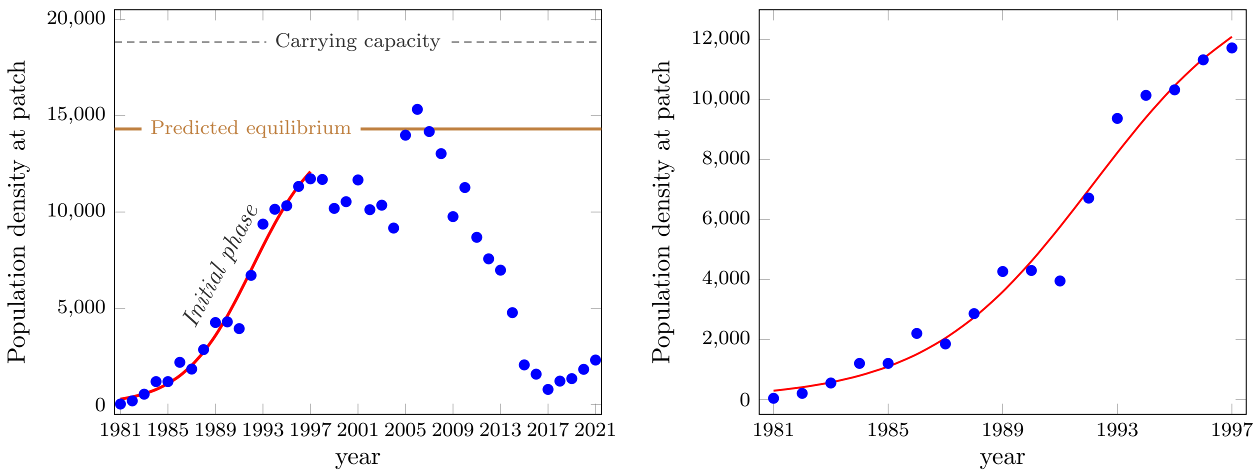


Figure S6. Left: Fitting of the initial phase (red line) from the establishment of the local population at La Banya since the arrival of the predators (1981–1997) using Equation (3.1) and the parameters displayed above. The horizontal dashed line shows the predicted carrying capacity, while the horizontal brown line shows the predicted equilibrium of the population.
Right: Zoom in the period 1981–1997.

Supplementary Section 4

Dynamics after the perturbation: dispersal by social copying

Here, we introduce and investigate the dynamics of the population with the model including linear dispersal and dispersal by social copying. The model reads:

$$\frac{\mathrm{d}}{\mathrm{d}t}x(t) = \varphi x(t) - \beta x(t)^2 - \lambda D(x(t), \mu, \sigma, \delta) \tag{4.1}$$

with the following parameters.

Parameter	Range or value	Ecological meaning or description
ε	0.11	Death rate estimated from field data [10]
α	0.3489494104672237	Population growth rate including death of individuals (without linear dispersal)
$\gamma = \alpha + \varepsilon$	0.4589494112062910	See Table 1 for details
$K = \frac{\gamma}{\beta}$	18822.79734	Carrying capacity
$x(0)$	$[0, K]$	Initial condition
ρ	$\mathbb{R}^+$	Linear (exponential) dispersal rate
$\varphi = \alpha - \rho$	$(-\infty, \alpha]$	Neat population growth rate including linear dispersal
β	$2.43826356697 \times 10^{-5}$	Intrinsic growth rate over the carrying capacity
Parameters concerning dispersal rate by social copying		
λ	$\mathbb{R}^+$	Dispersal rate
μ	$\mathbb{R}^+$	Tendency of dispersal function for small population sizes
σ	$\mathbb{R}^+$	Sharpness and smoothness of the dispersal function
δ	$\mathbb{R}^+$	Transition between small and large population sizes

Next, we introduce the chosen dispersal function for Model (2.1) able to mimic dispersal by social copying. We have chosen the so-called Elliot sigmoid function which includes different parameters which allow to obtain multitude of different shapes. The key point for choosing this function is that it typically increases at decreasing population values. Generically, the less the population at the patch, the largest value for this function and thus the higher dispersal rates. However, due to its plasticity, other behaviours can be found showing this increasing tendency at low population numbers in certain parts. For instance, some values of the function can decrease at low population values; for some other parameters one can obtain density-independent dispersal e.g., Figure S7 for $\sigma = 0.001$, see also Remark 4.2 and Figure S10); as well as different sigmoidal shapes.

4.1 Modelling dispersal by social copying

The nonlinear dispersal function that we propose for Model (4.1) (and model (2.1)) is given by:

$$D(x, \mu, \sigma, \delta) := \begin{cases} \frac{1-\mathcal{E}_{\mathsf{dir}}(x,\mu,\sigma,\delta)}{1-\mathcal{E}_{\mathsf{dir}}(0,\mu,\sigma,\delta)} & \text{when } 0 \le x \le \delta, \\ \frac{1-\mathcal{E}(x,\sigma,\delta)}{1-\mathcal{E}_{\mathsf{dir}}(0,\mu,\sigma,\delta)} & \text{when } x \ge \delta, \end{cases} \tag{4.2}$$

where

$$\mathcal{E}_{\mathsf{dir}}(x, \mu, \sigma, \delta) := \left(\mu \frac{\Theta + \sigma\delta}{2\Theta + \sigma\delta}\left(1 - \frac{x}{\delta}\right) + \frac{x}{\delta}\right)\mathcal{E}(x, \sigma, \delta), \tag{4.3}$$

and

$$\mathcal{E}(x, \sigma, \delta) := \frac{\sigma(x - \delta)}{\Theta + \sigma|x - \delta|}, \tag{4.4}$$

is an *Elliot sigmoid* Θ–scaled, σ–strengthened, and δ–displaced. All the parameters of the dispersal function are non-negative and we have fixed $\Theta := 1000$ (this parameter controls the scale in the independent variable x which is related with the order of magnitude of the carrying capacity K).

Below, we describe the meaning of the other parameters of the dispersal function by providing a brief mathematical description and displaying some examples of graphs for several illustrative sets of values of parameters. A more detailed and technical description of this function can be found in Section 4.1.1 below.

Proposition 4.1 (On the function $\boldsymbol{D(x, \mu, \sigma, \delta)}$). *For every $\mu, \delta \ge 0$ and $x \ge 0$ we have $D(x, \mu, 0, \delta) \equiv 1$. Moreover, for $\sigma > 0$ we have*

(a) The function $D(x, \mu, \sigma, \delta)$, as a function of x, is continuous, differentiable, and strictly positive.

(b) $D(0, \mu, \sigma, \delta) = 1$ and $\lim_{x\to+\infty} D(x, \mu, \sigma, \delta) = 0$.

(c) If $\mu \ge 1$, then $D(x, \mu, \sigma, \delta)$ is strictly decreasing as a function of x. Moreover, $\frac{\mathrm{d}}{\mathrm{d}x} D(x, \mu, \sigma, \delta)\big|_{x=0}$ is 0 when $\mu = 1$ and negative when $\mu > 1$.

(d) For $0 \le \mu < 1$ and $\delta > 0$, $D(x, \mu, \sigma, \delta)$ is a unimodal function with a maximum at $x^* \in (0, \delta)$ (that is, D is strictly increasing in $[0, x^*]$ and strictly decreasing in $[x^*, +\infty)$). In particular, $\frac{\mathrm{d}}{\mathrm{d}x} D(x, \mu, \sigma, \delta) > 0$ for every $x \in [0, x^*)$. On the other hand, for every $x \in \mathbb{R}^+$, $D(x, \mu, \sigma, \delta) \le D(x^*, \mu, \sigma, \delta) < 2$.

As described by Proposition 4.1, the function D is normalized to one at zero population size, and when the population size tends to infinity, the tendency to disperse converges to zero. Furthermore, D is designed so that the dispersal response of the population of birds generically increases when the population numbers at the patch diminish.

Parameter σ controls the sharpness of the jump and the smoothness of the dispersal function. In Figure S7 we display several graphs for several values of σ. On the left panel, for $\sigma \ge 1$, it can be observed that as σ goes to infinity the graph of D becomes less smooth and the transition from high to small values of the dispersal function is more quick and abrupt. On the right panel, for $\sigma < 1$, it is shown that as σ decreases, the dispersal function becomes flatter and as σ converges to zero the graph of D converges to the constant function one. This latter case corresponds to a constant i.e., independent of the size of the population, dispersal.

Parameter μ controls the tendency (derivative) of the dispersal function at population size $x = 0$. Indeed, for $\mu < 1$ the curve starts at $x = 0$ with increasing tendency; for $\mu = 0$ the curve starts with derivative zero; while for $\mu > 1$ the curve starts with decreasing tendency [see Proposition 4.1(c,d) and Figure S8]. Observe that the dispersal curve whenever $\mu \ge 1$ is globally strictly decreasing while for $\mu < 1$ it is unimodal. From a more ecological point of view, parameter μ controls how fast dispersal is initiated by the individuals after the ecological perturbation below a given population threshold (modelled by parameter δ). The reason for designing the dispersal function so that it is increasing for low population sizes and low values of μ is to allow the model to deal with a wider range of nonlinear behaviours.

Finally, the dispersal function has been built in such a way that for low population values the tendency to disperse is large while it becomes smaller for large population sizes. Indeed, this is how this function models dispersal by social copying. Parameter δ controls the transition between these two scenarios (see Figure S9).

4.1.1 Properties of the dispersal function $\boldsymbol{D(x, \mu, \sigma, \delta)}$

The goal of this sub-subsection is to prove Proposition 4.1, which summarizes the most relevant properties of the dispersal function $D(x, \mu, \sigma, \delta)$. To do this, we will first study the functions $\mathcal{E}(x, \sigma, \delta)$ and $\mathcal{E}_{\mathsf{dir}}(x, \mu, \sigma, \delta)$.

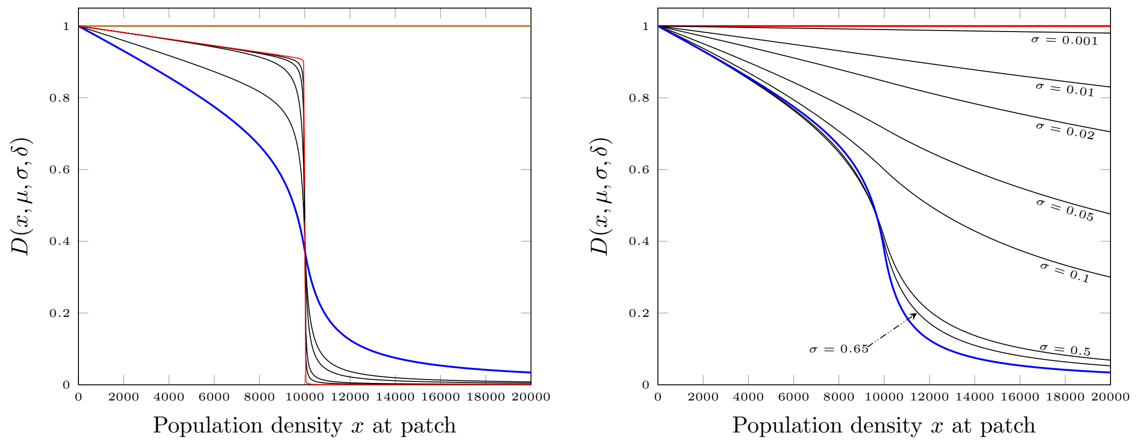


Figure S7. Shapes of the function $D(x,\mu,\sigma,\delta)$ used to model social copying behaviour during dispersal. We explore the ranges of $\sigma \geq 1$ (left panel) and $\sigma < 1$ (right panel) fixing $\delta = 10^4$.
Left: The brown graph corresponds to $\sigma = 0$, involving density-independent dispersal. The sigmoidal-like blue graph has been obtained with $(\mu,\sigma) = (2,1)$. The red curve, which is in some sense the limiting graph, is obtained with $(\mu = 1.2, \sigma = 10^3)$. The four black curves correspond, from bottom to top, to the following parameter values: $(\mu,\sigma) = (1.5,5)$, $(\mu,\sigma) = (1.2,10)$, $(\mu,\sigma) = (1.2,40)$, and $(\mu,\sigma) = (1.2,10^2)$.
Right: The red curve in this case is the limiting graph with $\sigma = 0$,The blue graph is the same than the one in the left panel. The black curves have been obtained fixing $\mu = 2$, and $\sigma = 0.65, 0.5, 0.1, 0.05, 0.02, 10^{-2}, 10{-}3$.

Remark 4.2 (Function $D(x,\mu,\sigma,\delta)$ reproduces density-independent dispersal for a range of parameters with positive volume)**.** There are two possible parameter combinations to get density-independence in function $D(x,\mu,\sigma,\delta)$. One possibility simply consists in setting $\sigma = 0$ (brown line in Figure S7(left)). The other possibility, shown in Figure S10 consists in taking $\mu \approx 1$ (e.g., $\mu \in [0.999, 1.001]$), σ large (e.g., $\sigma \geq 100$), and δ very large (e.g., $\sigma \geq 5 \times 10^6$). Observe that the range of parameters for the case $\sigma = 0$ has zero volume, while for the latter case has positive volume.

Lemma 4.3 (On the functions $\mathcal{E}(x,\sigma,\delta)$ and $\mathcal{E}_{\mathsf{dir}}(x,\mu,\sigma,\delta)$)**.** *For all $\mu,\sigma,\delta \geq 0$ and $x \geq 0$ we have*

(1) $\mathcal{E}(0,\sigma,\delta) = -\frac{\sigma\delta}{\Theta+\sigma\delta}$, *and* $\mathcal{E}_{\mathsf{dir}}(0,\mu,\sigma,\delta) = -\mu\frac{\sigma\delta}{2\Theta+\sigma\delta}$,

(2) $\mathcal{E}_{\mathsf{dir}}(\delta,\mu,\sigma,\delta) = \mathcal{E}(\delta,\sigma,\delta) = 0$,

(3) $\mathcal{E}_{\mathsf{dir}}(x,\mu,0,\delta) = \mathcal{E}(x,0,\delta) \equiv 0$ *for every* $x \geq 0$,

(4) $-1 < \mathcal{E}(x,\sigma,\delta) < 1$,

(5) $\frac{\mathrm{d}}{\mathrm{d}x}\mathcal{E}(x,\sigma,\delta) = \frac{\Theta\sigma}{(\Theta+\sigma|x-\delta|)^2} > 0$, *and*

(6) $\lim_{x\to+\infty}\mathcal{E}(x,\sigma,\delta) = 1$ *provided that* $\sigma > 0$.

When $\sigma > 0$, $\mathcal{E}$ and $\mathcal{E}_{\mathsf{dir}}$ are continuous as functions of x. Moreover, for $\mu \geq 0$ and $0 \leq x \leq \delta$,

$$\frac{\mathrm{d}}{\mathrm{d}x}\mathcal{E}_{\mathsf{dir}}(x,\mu,\sigma,\delta) = \frac{\sigma}{\delta(2\Theta+\sigma\delta)(\Theta+\sigma z)}\left(-\Gamma z + \Big(\mu\delta(\Theta+\sigma\delta) + \Gamma(\delta - z)\Big)\frac{\Theta}{\Theta+\sigma z}\right),$$

where $\Gamma := (2-\mu)\Theta + (1-\mu)\sigma\delta$ *and* $z = \delta - x$.

Proof. Statements (1–6) are obtained by direct computation. The fact that when $\sigma > 0$, $\mathcal{E}$ and $\mathcal{E}_{\mathsf{dir}}$ are continuous follows easily from the definitions of $\mathcal{E}$ and $\mathcal{E}_{\mathsf{dir}}$. Now we prove the last statement of the lemma.

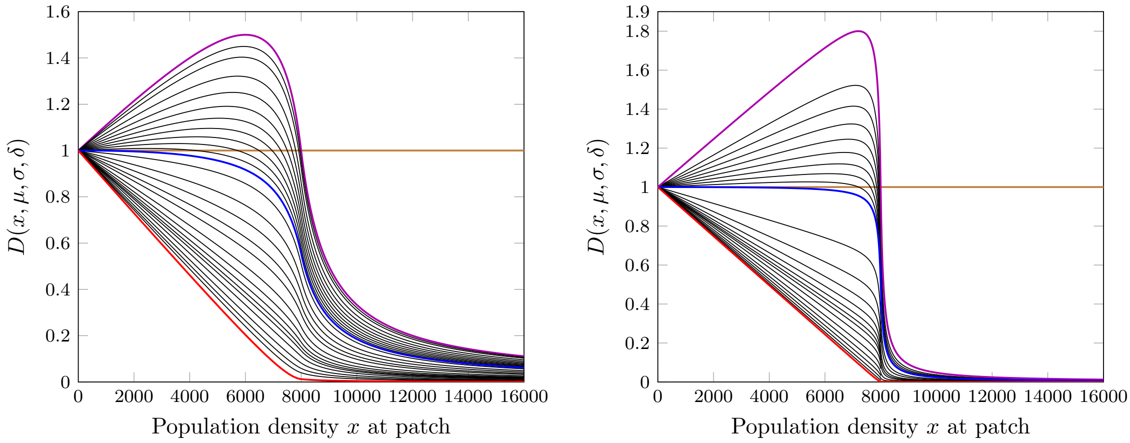


Figure S8. Shapes of the dispersal function $D(x, \mu, \sigma, \delta)$ for $\delta = 8 \times 10^3$ and (left picture) $\sigma = 1$ and (right picture) $\sigma = 10$. The violet and the blue curves correspond to $\mu = 0$ and $\mu = 1$, respectively. The red graph is, in some sense, the limiting case: $\mu = 100$ for the left panel and $\mu = 500$ for the right panel. All black curves are organised, from top to bottom, by increasing value of μ. Thus, all black curves between the violet and the blue curves correspond to $\mu < 1$ while the blue curves between the blue and red curves are obtained for $\mu > 1$.

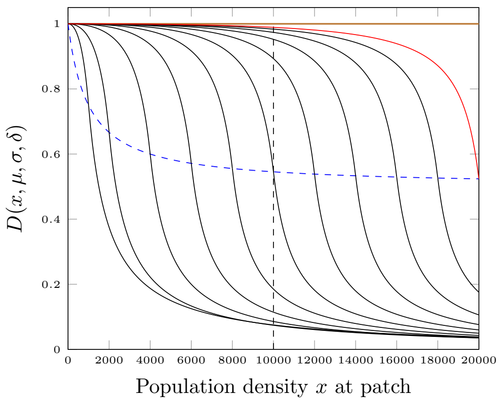


Figure S9. Shapes of the dispersal function $D(x, \mu, \sigma, \delta)$ for $\sigma = \mu = 1$ and different values of δ. Each curve is obtained by using the value of δ given by the x coordinate of the intersection of the blue dashed line with the curve. The vertical dashed line identifies the curve obtained with $\delta = 10^4$.

From the definition of $\mathcal{E}_{\mathsf{dir}}$ and (5) we have:

$$\begin{aligned}
\frac{\mathrm{d}}{\mathrm{d}x}\,\mathcal{E}_{\mathsf{dir}}(x,\mu,\sigma,\delta) &= \frac{1}{\delta}\left(1 - \frac{\mu(\Theta+\sigma\delta)}{2\Theta+\sigma\delta}\right)\mathcal{E}(x,\sigma,\delta) + \\
&\qquad \left(\mu\frac{\Theta+\sigma\delta}{2\Theta+\sigma\delta} + \frac{x}{\delta}\left(1 - \frac{\mu(\Theta+\sigma\delta)}{2\Theta+\sigma\delta}\right)\right)\frac{\mathrm{d}}{\mathrm{d}x}\,\mathcal{E}(x,\sigma,\delta) \\
&= \frac{1}{\delta}\frac{\Gamma}{2\Theta+\sigma\delta}\frac{\sigma(x-\delta)}{\Theta+\sigma(\delta-x)} + \\
&\qquad \left(\mu\frac{\Theta+\sigma\delta}{2\Theta+\sigma\delta} + \frac{\Gamma}{\delta(2\Theta+\sigma\delta)}x\right)\frac{\Theta\sigma}{(\Theta+\sigma(\delta-x))^2} \\
&= \frac{\sigma}{\delta(2\Theta+\sigma\delta)(\Theta+\sigma z)}\left(-\Gamma z + \Big(\mu\delta(\Theta+\sigma\delta) + \Gamma(\delta-z)\Big)\frac{\Theta}{\Theta+\sigma z}\right).
\end{aligned}$$

□

We are ready to prove Proposition 4.1.

Proof of Proposition 4.1. The fact that $D(x, \mu, 0, \delta) \equiv 1$ and Statement (b) follow from the definition of D and Lemma 4.3(3,6). Now we assume $\sigma > 0$. The continuity of D as a function of x follows from its definition

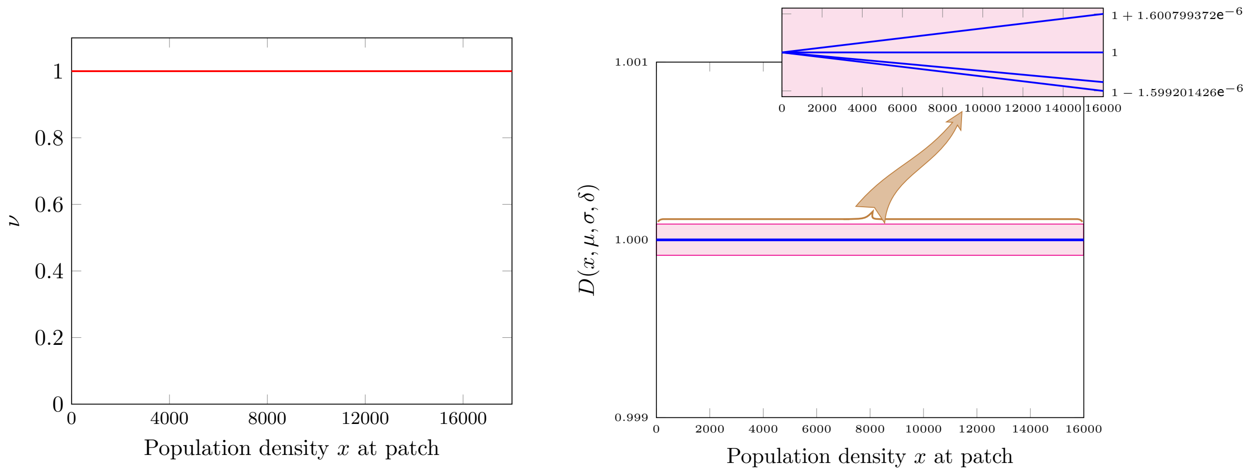


Figure S10. Left: Density-independent dispersal for $\sigma = 0$ and arbitrary values of $\mu, \delta \geq 0$.
Right: Density-independent dispersal using function D for different range of parameters with e.g., $\mu \in [0.99, 1.001]$, $\sigma > 10^3$, and δ large. The inset displayed with the arrow shows an enlarged view close to $D(x, \mu, \sigma, \delta) = 1$

and Lemma 4.3. Moreover, by Lemma 4.3(1),

$$D(x,\mu,\sigma,\delta) = \frac{1}{1+\frac{\mu\sigma\delta}{2\Theta+\sigma\delta}} \begin{Bmatrix} \big(1-\mathcal{E}(x,\sigma,\delta)\big) & \text{when } x \geq \delta, \\ \big(1-\mathcal{E}_{\mathsf{dir}}(x,\mu,\sigma,\delta)\big) & \text{when } 0 \leq x \leq \delta, \end{Bmatrix} = \frac{2\Theta+\sigma\delta}{2\Theta+(1+\mu)\sigma\delta} \begin{cases} \big(1-\mathcal{E}(x,\sigma,\delta)\big) & \text{when } x \geq \delta, \\ \big(1-\mathcal{E}_{\mathsf{dir}}(x,\mu,\sigma,\delta)\big) & \text{when } 0 \leq x \leq \delta. \end{cases} \tag{4.5}$$

Observe that $\frac{2\Theta+\sigma\delta}{2\Theta+(1+\mu)\sigma\delta} > 0$ and, by Lemma 4.3(4), $1-\mathcal{E}(x,\sigma,\delta) > 0$ for $x \geq \delta$. Moreover, for $0 \leq x \leq \delta$,

$$\mu\frac{\Theta+\sigma\delta}{2\Theta+\sigma\delta}\left(1-\frac{x}{\delta}\right)+\frac{x}{\delta} \geq 0 \tag{4.6}$$

and, by Lemma 4.3(2,5), $\mathcal{E}(x,\sigma,\delta) \leq 0$. Consequently, for $0 \leq x \leq \delta$,

$$1-\mathcal{E}_{\mathsf{dir}}(x,\mu,\sigma,\delta) = 1-\left(\mu\frac{\Theta+\sigma\delta}{2\Theta+\sigma\delta}\left(1-\frac{x}{\delta}\right)+\frac{x}{\delta}\right)\mathcal{E}(x,\sigma,\delta) \geq 1 \tag{4.7}$$

Thus, $D(x,\mu,\sigma,\delta)$ is strictly positive.

For $\sigma, \delta > 0$, from the definition of $\mathcal{E}_{\mathsf{dir}}$ and Lemma 4.3(2) we get,

$$\begin{aligned} \frac{\mathrm{d}}{\mathrm{d}x}\,\mathcal{E}_{\mathsf{dir}}(x,\mu,\sigma,\delta)\big|_{x=\delta} &= \frac{\mathrm{d}}{\mathrm{d}x}\left(\mu\frac{\Theta+\sigma\delta}{2\Theta+\sigma\delta}\left(1-\frac{x}{\delta}\right)+\frac{x}{\delta}\right)\Bigg|_{x=\delta}\mathcal{E}(\delta,\sigma,\delta) + \\ &\quad \left(\mu\frac{\Theta+\sigma\delta}{2\Theta+\sigma\delta}\left(1-\frac{x}{\delta}\right)+\frac{x}{\delta}\right)\Bigg|_{x=\delta}\frac{\mathrm{d}}{\mathrm{d}x}\,\mathcal{E}(x,\sigma,\delta)\big|_{x=\delta} = \frac{\mathrm{d}}{\mathrm{d}x}\,\mathcal{E}(x,\sigma,\delta)\big|_{x=\delta}. \end{aligned}$$

Hence, by Equation (4.5), $D(x,\mu,\sigma,\delta)$ is differentiable as a function of $x \geq 0$, and (a) holds.

To prove (c) and (d) we need to study the monotonicity of the map D. For $\sigma > 0$ and $x \geq \delta$, $D(x,\mu,\sigma,\delta)$ is strictly decreasing by Equation (4.5) and Lemma 4.3(5).

On the other hand, for $\sigma, \delta > 0$ and $0 \leq x \leq \delta$ we have by Lemma 4.3 and Equation (4.5),

$$\begin{aligned} \frac{\mathrm{d}}{\mathrm{d}x}D(x,\mu,\sigma,\delta) &= -\frac{2\Theta+\sigma\delta}{2\Theta+(1+\mu)\sigma\delta}\,\frac{\sigma}{\delta(2\Theta+\sigma\delta)(\Theta+\sigma z)}\left(-\Gamma z+\Big(\mu\delta(\Theta+\sigma\delta)+\Gamma(\delta-z)\Big)\frac{\Theta}{\Theta+\sigma z}\right) \\ &= \frac{\sigma}{\delta(\Theta+\sigma z)^2(2\Theta+(1+\mu)\sigma\delta)}\left(\Gamma z(\Theta+\sigma z)-\Theta\Big(\mu\delta(\Theta+\sigma\delta)+\Gamma(\delta-z)\Big)\right) \\ &= \frac{\sigma}{\delta(\Theta+\sigma z)^2(2\Theta+(1+\mu)\sigma\delta)}\Big(\Gamma\big(\Theta(2z-\delta)+\sigma z^2\big)-\mu\Theta\delta(\Theta+\sigma\delta)\Big). \end{aligned}$$

Moreover, since $\Theta, \sigma, \delta > 0$ and $\mu \geq 0$, we have

$$\frac{\sigma}{\delta(\Theta+\sigma z)^2(2\Theta+(1+\mu)\sigma\delta)} > 0.$$

Since $z=\delta$ whenever $x=0$, from the above expression for $\frac{\mathrm{d}}{\mathrm{d}x} D(x,\mu,\sigma,\delta)$ we obtain,

$$\begin{aligned}
\frac{\mathrm{d}}{\mathrm{d}x} D(x,\mu,\sigma,\delta)\big|_{x=0} &= \frac{\sigma}{\delta(\Theta+\sigma\delta)^2(2\Theta+(1+\mu)\sigma\delta)}\Big(\Gamma\big(\Theta\delta+\sigma\delta^2\big)-\mu\Theta\delta(\Theta+\sigma\delta)\Big)\\
&= \frac{\sigma}{(\Theta+\sigma\delta)(2\Theta+(1+\mu)\sigma\delta)}\Big(\Gamma-\mu\Theta\Big)\\
&= \frac{\sigma}{(\Theta+\sigma\delta)(2\Theta+(1+\mu)\sigma\delta)}\Big((2-\mu)\Theta+(1-\mu)\sigma\delta-\mu\Theta\Big)\\
&= \frac{\sigma}{(\Theta+\sigma\delta)(2\Theta+(1+\mu)\sigma\delta)}(1-\mu)(2\Theta+\sigma\delta).
\end{aligned}$$

Since $2\Theta+\sigma\delta>0$ it follows that $\frac{\mathrm{d}}{\mathrm{d}x} D(x,\mu,\sigma,\delta)\big|_{x=0}$ is positive when $0\le\mu<1$, zero when $\mu=1$, and negative when $\mu>1$.

Now we study the monotonicity of $D(x,\mu,\sigma,\delta)$ for $0<x\le\delta$ (which is equivalent to $0\le z<\delta$). When $\mu\ge 1$ we have $\Gamma\le\Theta$, and hence

$$\begin{aligned}
\Gamma(\Theta(2z-\delta)+\sigma z^2)-\mu\Theta\delta(\Theta+\sigma\delta) &\le \Theta(\Theta(2z-\delta)+\sigma z^2)-\Theta\delta(\Theta+\sigma\delta)\\
&= \Theta\big(2\Theta(z-\delta)+\sigma(z^2-\delta^2)\big)<0.
\end{aligned}$$

Thus, in summary, $\frac{\mathrm{d}}{\mathrm{d}x} D(x,\mu,\sigma,\delta)<0$ for $0<x\le\delta$. Since we already know that $D(x,\mu,\sigma,\delta)\big|_{[\delta,+\infty)}$ is strictly decreasing, it follows that $D(x,\mu,\sigma,\delta)$ is globally strictly decreasing as a function of $x\in\mathbb{R}^+$, and (c) is proved.

Next, to prove (d), we study the shape of $D(x,\mu,\sigma,\delta)\big|_{[0,\delta]}$ when $0\le\mu<1$. To do this notice that $\Gamma(\Theta(2z-\delta)+\sigma z^2)-\mu\Theta\delta(\Theta+\sigma\delta)=0$ is equivalent to

$$\Gamma(\Theta(2z-\delta)+\sigma z^2)=\mu\Theta\delta(\Theta+\sigma\delta)$$

which, in turn, is equivalent to

$$z(2\Theta+\sigma z)-\Theta\delta=\Theta(2z-\delta)+\sigma z^2=\frac{\mu\Theta\delta(\Theta+\sigma\delta)}{\Gamma},$$

and to

$$z(2\Theta+\sigma z)=\Theta\delta+\frac{\mu\Theta\delta(\Theta+\sigma\delta)}{\Gamma}.$$

Now observe that, for $0\le\mu<1$ we have $\Theta<\Gamma$, and hence

$$z(2\Theta+\sigma z)\Big|_{z=0}=0<\Theta\delta+\mu\frac{\Theta}{\Gamma}\delta(\Theta+\sigma\delta)<\delta(2\Theta+\sigma\delta)=z(2\Theta+\sigma z)\Big|_{z=\delta}.$$

Consequently, since $z\mapsto z(2\Theta+\sigma z)$ is a continuous strictly increasing function of $z\ge 0$, there exists a unique $x^*(\mu)=\delta-z^*(\mu)\in(0,\delta)$ such that

$$(\delta-x^*(\mu))(2\Theta+\sigma(\delta-x^*(\mu)))=\Theta\delta+\frac{\mu\Theta\delta(\Theta+\sigma\delta)}{\Gamma},$$

which is equivalent to

$$\Gamma(\Theta(2(\delta-x^*(\mu))-\delta)+\sigma(\delta-x^*(\mu))^2)-\mu\Theta\delta(\Theta+\sigma\delta)=0,$$

and to $\frac{\mathrm{d}}{\mathrm{d}x} D(x^*(\mu),\mu,\sigma,\delta)=0$. Then, the unicity of the critical point x^* and the fact that $\frac{\mathrm{d}}{\mathrm{d}x} D(x,\mu,\sigma,\delta)\Big|_{x=0}$ is positive when $0\le\mu<1$ tells us that $D(x,\mu,\sigma,\delta)\Big|_{[0,\delta]}$ is a unimodal map with $D(x,\mu,\sigma,\delta)\Big|_{[0,x^*]}$ strictly increasing and $D(x,\mu,\sigma,\delta)\Big|_{[x^*,\delta]}$ strictly decreasing. By using again the fact that $D(x,\mu,\sigma,\delta)\big|_{[\delta,+\infty)}$ is strictly decreasing, we get (d) except for the fact that $D(x^*,\mu,\sigma,\delta)<2$. To prove it observe that, for $0\le x\le\delta$, in view of Lemma 4.3(5,1) and (4.6) and (4.7) we have

$$
\begin{aligned}
D(x,\mu,\sigma,\delta) &= \frac{2\Theta+\sigma\delta}{2\Theta+(1+\mu)\sigma\delta}\big(1-\mathcal{E}_{\mathsf{dir}}(x,\mu,\sigma,\delta)\big) \le 1-\mathcal{E}_{\mathsf{dir}}(x,\mu,\sigma,\delta)\\
&= 1-\left(\mu\frac{\Theta+\sigma\delta}{2\Theta+\sigma\delta}\left(1-\frac{x}{\delta}\right)+\frac{x}{\delta}\right)\mathcal{E}(x,\sigma,\delta)\\
&\le 1+\left(\mu\frac{\Theta+\sigma\delta}{2\Theta+\sigma\delta}\left(1-\frac{x}{\delta}\right)+\frac{x}{\delta}\right)\big(-\mathcal{E}(0,\sigma,\delta)\big) \le 1+\left(\frac{\Theta+\sigma\delta}{2\Theta+\sigma\delta}\left(1-\frac{x}{\delta}\right)+\frac{x}{\delta}\right)\frac{\sigma\delta}{\Theta+\sigma\delta}\\
&= 1+\frac{\sigma\delta}{2\Theta+\sigma\delta}+\frac{x}{\delta}\frac{\sigma\delta}{\Theta+\sigma\delta}\left(1-\frac{\Theta+\sigma\delta}{2\Theta+\sigma\delta}\right) = 1+\frac{\sigma\delta}{2\Theta+\sigma\delta}+\frac{x}{\delta}\frac{\sigma\delta}{\Theta+\sigma\delta}\frac{\Theta}{2\Theta+\sigma\delta}\\
&< 1+\frac{\Theta+\sigma\delta}{2\Theta+\sigma\delta} < 2
\end{aligned}
$$

because $1-\frac{x}{\delta}\ge 0$. $\square$

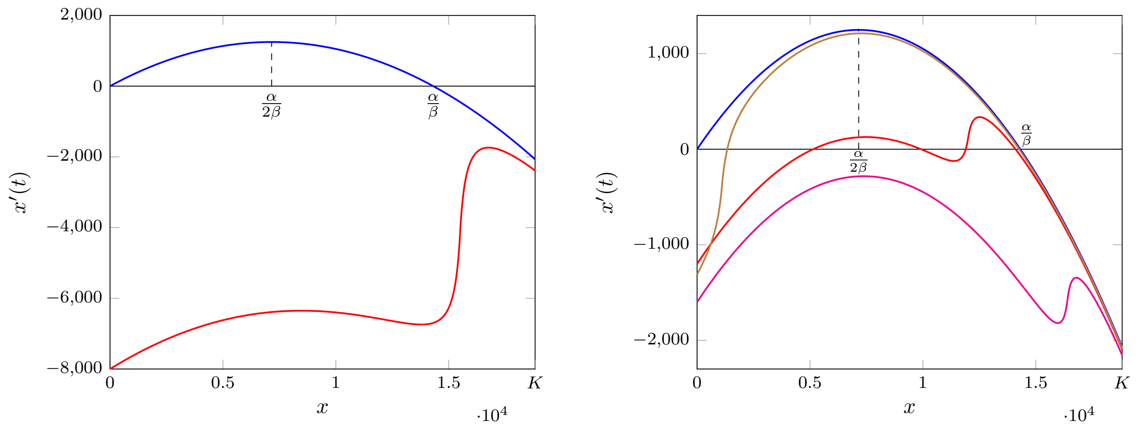


Figure S11. The vector field (4.1) in the x-interval $[0,K]$, in parameter's realistic cases. In blue it is shown the graph of $f(x)$ and in red (brown and magenta) several possible graphs of $F(x)$. In all cases, the values of K, α and β are the ones obtained as population's characteristics by fitting the data to the initial phase (see the table in Page 15). Likewise, for the red, brown and magenta curves, the values $\mu = 1.2$ and $\sigma = 3.2$ are fixed. The additional parameters take the following values:

- For the red curve in the left picture: $\delta = 15500$, and $\lambda = 8000$.
- For the brown curve in the right picture: $\delta = 1100$, and $\lambda = 1310$.
- For the red curve in the right picture: $\delta = 12000$, and $\lambda = 1200$.
- For the magenta curve in the right picture: $\delta = 16400$, and $\lambda = 1600$.

Summing up the previous results in Section S 4 we know that, for $\sigma > 0$ and $\mu \ge 1$, $D(x,\mu,\sigma,\delta)$ is a continuous, positive, strictly decreasing function such that $D(0,\mu,\sigma,\delta) = 1$. Hence the function $-\lambda D(x,\mu,\sigma,\delta)$ is strictly increasing as a function of x and since, by Lemma 3.2, we know that $f\big|_{\left(0,\frac{\alpha}{2\beta}\right]}$ is strictly increasing we easily get (see Figures S11 and S12):

Lemma 4.4. *The function $F(x) = x(\alpha-\beta x) - \lambda D(x,\mu,\sigma,\delta) = f(x) - \lambda D(x,\mu,\sigma,\delta)$ verifies*

$$
\begin{aligned}
F(0) &= f(0) - \lambda = -\lambda < 0,\\
F(K) &= f(K) - \lambda D(K,\mu,\sigma,\delta) = -K\varepsilon - \lambda D(K,\mu,\sigma,\delta) < 0,
\end{aligned}
$$

and

$$x(\alpha-\beta x) - \lambda \le F(x) \le x(\alpha-\beta x) - \lambda D(K,\mu,\sigma,\delta) < x(\alpha-\beta x)$$

for every $x \in [0,K]$. Moreover, $F\big|_{\left(0,\frac{\alpha}{2\beta}\right]}$ is strictly increasing.

Next we study the full shape of the function $F\big|_{[0,K]}$ (see Figures S11 and S12).

Lemma 4.5 (On the shape of $F\big|_{[0,K]}$)**.** *The function $F\big|_{[0,K]}$ verifies one of the following statements:*

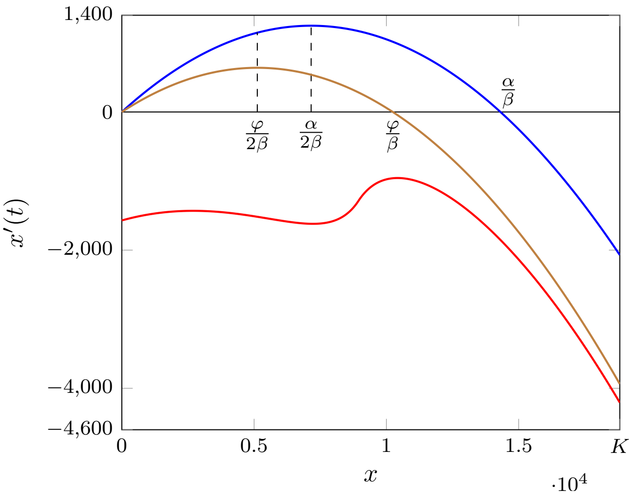


Figure S12. The vector field (4.1) in the interval $[0, K]$. In blue it is shown the graph of $f(x)$, in brown it is shown the graph of $x \cdot (\varphi - \beta x)$, and in red the graph of $F(x)$. The parameter values used in this picture are the following: K, α and β take the values from the table in Page 15, as before. For the red and brown curves we take the parameter's values corresponding to the best fit with the phase 2006–2017 data. These are: $\varphi = 0.2497248909716255$, $\lambda = 1570.2313809039706030$, $\mu = 0$, $\sigma = 0.4904756364357690$, and $\delta = 8944.2282749675759987$.

(A) F has at most one critical point in the interval $[0, K]$, and this critical point is an inflexion point. Hence, $F\big|_{[0,K]}$ is strictly increasing.

(B) F has a unique critical point c in the interval $[0, K]$. The critical point c is a maximum and belongs to $\left(\frac{\alpha}{2\beta}, K\right]$. Hence, $F\big|_{[0,c]}$ is strictly increasing and, when $c < K$, $F\big|_{[c,K]}$ is strictly decreasing.

(C) F has at most two critical points in the interval $[0, K]$, and both critical points belong to the interval $\left(\frac{\alpha}{2\beta}, K\right)$. One of them, donoted by c, is a maximum and the other one is an inflexion point. Hence, $F\big|_{[0,c]}$ is strictly increasing and $F\big|_{[c,K]}$ is strictly decreasing.

(D) F has exactly two critical points $\frac{\alpha}{2\beta} < c^+ < c^- \le K$ in the interval $[0, K]$. c^+ is a maximum while c^- is a minimum. Hence, $F\big|_{[0,c^+]}$ and $F\big|_{[c^-,K]}$ (when $c^- < K$) are strictly increasing while $F\big|_{[c^+,c^-]}$ is strictly decreasing.

(E) F has exactly three critical points $\frac{\alpha}{2\beta} < c_1^+ < c^- < c_2^+ < K$ in the interval $[0, K]$. c_1^+ and c_2^+ are maxima while c^- is a minimum. Hence, $F\big|_{[0,c_1^+]}$ $F\big|_{[c^-,c_2^+]}$ are strictly increasing while $F\big|_{[c_1^+,c^-]}$ and $F\big|_{[c_2^+,K]}$ are strictly decreasing.

Proof. In this proof we will use the expressions for $\frac{d}{dx}D(x, \mu, \sigma, \delta)$ from above. The whole proof amounts to control the zeros of

$$\begin{aligned}
F'(x) &= f'(x) \\
&= \alpha - 2\beta x + \frac{d}{dx}\left(-M + M \begin{cases} \mathcal{E}_{\mathsf{dir}}(x, \mu, \sigma, \delta) & \text{when } 0 \le x \le \delta, \\ \mathcal{E}(x, \sigma, \delta) & \text{when } x \ge \delta, \end{cases}\right) \\
&= \alpha - 2\beta x + M \begin{cases} \frac{1}{\delta(2\Theta+\sigma\delta)} \frac{A - z\sigma\Gamma\left(2\Theta+\sigma z\right)}{(\Theta+\sigma z)^2} & \text{when } 0 \le x \le \delta, \\ \frac{\Theta\sigma}{(\Theta+\sigma(x-\delta))^2} & \text{when } x \ge \delta, \end{cases}
\end{aligned}$$

where $z = \delta - x$ and

$$\begin{aligned}
M &:= \frac{\lambda}{1 - \mathcal{E}_{\mathsf{dir}}(0, \mu, \sigma, \delta)} > 0 \text{ (recall that } \lambda > 0 \text{ and } 1 - \mathcal{E}_{\mathsf{dir}}(0, \mu, \sigma, \delta) > 0), \\
\Gamma &:= (2 - \mu)\Theta + (1 - \mu)\sigma\delta = (2\Theta + \sigma\delta) - \mu(\Theta + \sigma\delta), \text{ and} \\
A &:= \Theta\sigma\delta\big(\mu(\Theta + \sigma\delta) + \Gamma\big) = \Theta\sigma\delta\big(2\Theta + \sigma\delta\big) > 0.
\end{aligned}$$

The expression,

$$-\lambda\frac{d}{dx}D(x,\mu,\sigma,\delta) = M\begin{cases}\frac{1}{\delta(2\Theta+\sigma\delta)}\frac{A-z\sigma\Gamma\big(2\Theta+\sigma z\big)}{(\Theta+\sigma z)^2} & \text{when } 0\le x\le\delta,\\ \frac{\Theta\sigma}{(\Theta+\sigma(x-\delta))^2} & \text{when } x\ge\delta,\end{cases} \tag{4.8}$$

is strictly positive because $-\lambda D(x,\mu,\sigma,\delta)$ is strictly increasing as a function of x. Therefore, $F'(x)>0$ for every $x\in\big[0,\frac{\alpha}{2\beta}\big]$ (see Lemma 4.4).

So, if F has a critical point at $x\in[0,K]$, then $x>\frac{\alpha}{2\beta}$ and $F'(x)=0$, which is equivalent to

$$2\beta x-\alpha = M\begin{cases}\frac{1}{\delta(2\Theta+\sigma\delta)}\frac{A-z\sigma\Gamma\big(2\Theta+\sigma z\big)}{(\Theta+\sigma z)^2} & \text{when } 0\le x\le\delta,\\ \frac{\Theta\sigma}{(\Theta+\sigma(x-\delta))^2} & \text{when } x\ge\delta.\end{cases} \tag{4.9}$$

Concerning the monotonicity properties of (4.8) we have:

$$\begin{aligned}\frac{d}{dz}\frac{A-z\sigma\Gamma\big(2\Theta+\sigma z\big)}{(\Theta+\sigma z)^2} &= -\frac{2\sigma(\Gamma\Theta^2+A)}{(\Theta+\sigma z)^3} = \\ &-\frac{2\sigma\Theta\big(\Theta(2\Theta+\sigma\delta)-\mu\Theta(\Theta+\sigma\delta)+\sigma\delta(2\Theta+\sigma\delta)\big)}{(\Theta+\sigma z)^3} = \\ &-\frac{2\sigma\Theta\big((2\Theta+\sigma\delta)(\Theta+\sigma\delta)-\mu\Theta(\Theta+\sigma\delta)\big)}{(\Theta+\sigma z)^3} = \\ &-\frac{2\sigma\Theta(\Theta+\sigma\delta)\big(2\Theta+\sigma\delta-\mu\Theta\big)}{(\Theta+\sigma z)^3}.\end{aligned}$$

So, since $M, \frac{1}{\delta(2\Theta+\sigma\delta)}>0$ we see that

$$-\lambda\left(\frac{d}{dx}D(x,\mu,\sigma,\delta)\right)\Bigg|_{[0,\delta]} = M\frac{1}{\delta(2\Theta+\sigma\delta)}\frac{A-(\delta-x)\sigma\Gamma\big(2\Theta+\sigma(\delta-x)\big)}{(\Theta+\sigma(\delta-x))^2}\Bigg|_{[0,\delta]},$$

as a function of x, is strictly increasing when $\mu<2+\frac{\sigma\delta}{\Theta}$, constant when $\mu=2+\frac{\sigma\delta}{\Theta}$, and strictly decreasing otherwise (when $\mu>2+\frac{\sigma\delta}{\Theta}$).

On the other hand,

$$-\lambda\left(\frac{d}{dx}D(x,\mu,\sigma,\delta)\right)\Bigg|_{[\delta,+\infty)} = M\frac{\Theta\sigma}{(\Theta+\sigma(x-\delta))^2}$$

is strictly decreasing as a function of $x\ge\delta$ because $M>0$.

Summarising, the right hand side of Equation (4.9) restricted to the interval $\big[\frac{\alpha}{2\beta},K\big]$ is strictly positive and (see Figure S13):

(A) decreasing when either $\delta\le\frac{\alpha}{2\beta}$ or $\mu\ge 2+\frac{\sigma\delta}{\Theta}$,

(B) strictly increasing when $\delta\ge K$ and $\mu<2+\frac{\sigma\delta}{\Theta}$, and

(C) strictly increasing on $\big[\frac{\alpha}{2\beta},\delta\big]$ and strictly decreasing on $[\delta,K]$, when $\delta\in\big(\frac{\alpha}{2\beta},K\big)$ and $\mu<2+\frac{\sigma\delta}{\Theta}$.

Observe also that $-f'(x)=2\beta x-\alpha$ is affine with positive slope and vanishes at $x=\frac{\alpha}{2\beta}$. So,

$$(2\beta x-\alpha)\big|_{x=\frac{\alpha}{2\beta}} = 0 < -\lambda\big(\frac{d}{dx}D(x,\mu,\sigma,\delta)\big)\big|_{x=\frac{\alpha}{2\beta}}. \tag{4.10}$$

Consequently, the number κ of intersection points of the curves $-f'(x)=2\beta x-\alpha$ and $-\lambda\frac{d}{dx}D(x,\mu,\sigma,\delta)$ in the interval $\big(\frac{\alpha}{2\beta},K\big]$ verifies (see again Figure S13):

(A) $\kappa\le 1$ when either $\delta\le\frac{\alpha}{2\beta}$ or $\mu\ge 2+\frac{\sigma\delta}{\Theta}$. Moreover, in this case $\kappa=1$ if and only if

$$-\lambda\big(\frac{d}{dx}D(x,\mu,\sigma,\delta)\big)\big|_{x=K}\le 2\beta K-\alpha.$$

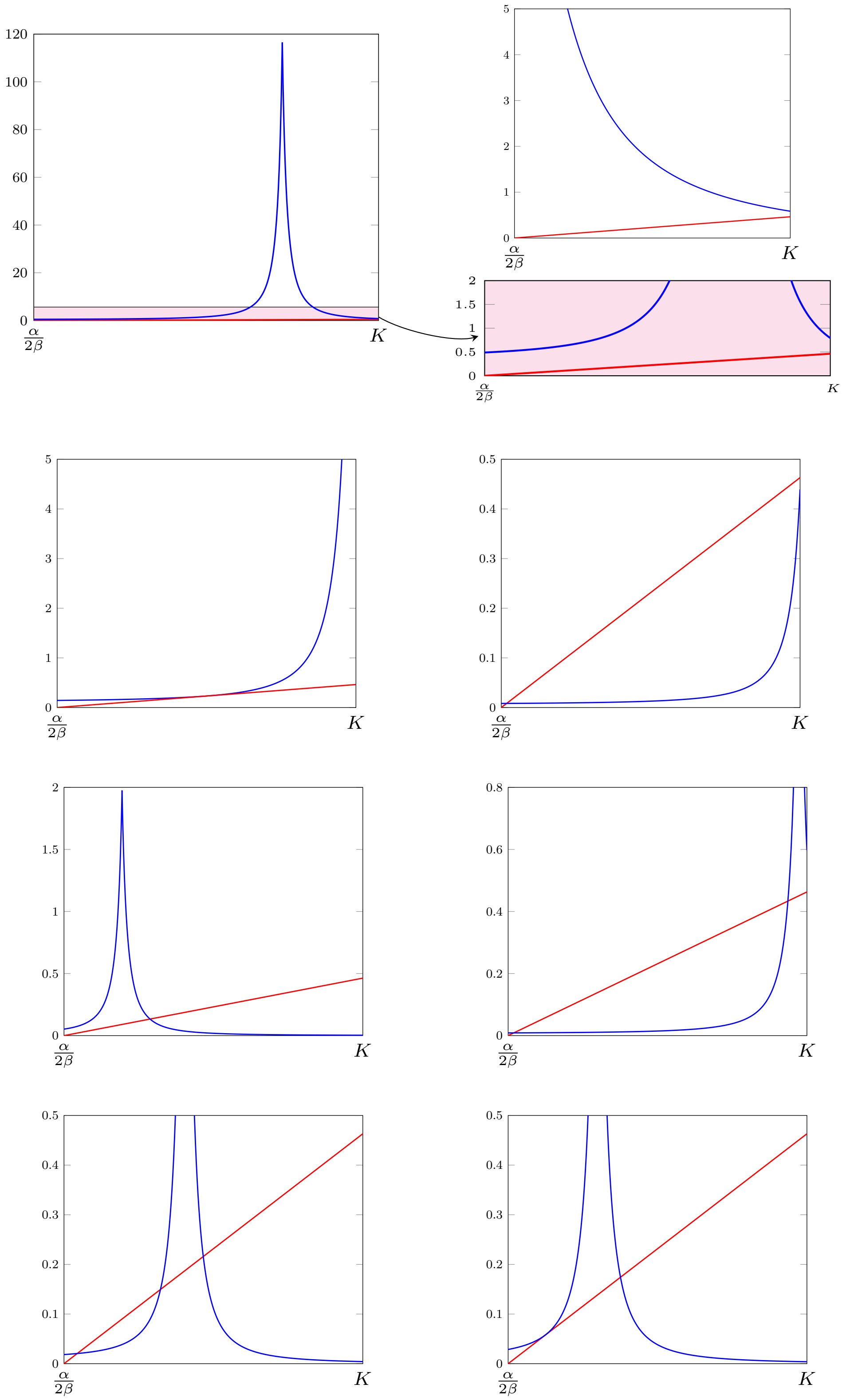


Figure S13. The graphs of $-f'(x) = 2\beta x - \alpha$ (in red) for the estimated values of α and β, and the right hand side of Equation (4.9) (in blue) for several realistic values of the parameters $\lambda > 0$, $\mu \geq 1$; $\sigma > 0$ and $\delta > 0$.

(B) When $\delta \geq K$ and $\mu < 2 + \frac{\sigma\delta}{\Theta}$ we have $\kappa \in \{0, 1, 2\}$.
Indeed, if $-\lambda\big(\frac{d}{dx}D(x, \mu, \sigma, \delta)\big) \geq 2\beta x - \alpha$ for every $x \in \big(\frac{\alpha}{2\beta}, K\big]$, then $\kappa \in \{0, 1\}$. On the other hand, if $-\lambda\big(\frac{d}{dx}D(x, \mu, \sigma, \delta)\big)\big|_{x=K} < 2\beta K - \alpha$ then $\kappa = 1$ (see the two pictures in the second row of Figure S13). The remaining case is when

$$\begin{aligned} &-\lambda\Big(\frac{d}{dx}D(x, \mu, \sigma, \delta)\Big)\Big|_{x=K} \geq 2\beta K - \alpha \text{ and, simultaneously,} \\ &-\lambda\Big(\frac{d}{dx}D(x, \mu, \sigma, \delta)\Big)\Big|_{x=y} < 2\beta y - \alpha \text{ for some } y \in \big(\tfrac{\alpha}{2\beta}, K\big). \end{aligned}$$

Clearly we have $F'(y) < 0$ and, in view of (4.10), $F'\big(\frac{\alpha}{2\beta}\big) > 0$. Thus, there exists $\widetilde{y} \in \big(\frac{\alpha}{2\beta}, y\big)$ such that $F'(\widetilde{y}) = 0$ and $F'(x) > 0$ for every $x \in \big(\frac{\alpha}{2\beta}, \widetilde{y}\big)$. On the other hand, observe that

$$F''(x)\big|_{\left[\frac{\alpha}{2\beta}, K\right]} = M\frac{1}{\delta(2\Theta + \sigma\delta)}\frac{2\sigma\Theta(\Theta + \sigma\delta)\big(2\Theta + \sigma\delta - \mu\Theta\big)}{(\Theta + \sigma z)^3} - 2\beta, \tag{4.11}$$

is strictly increasing as a function of x. Hence, it must be that $F'' < 0$ in the interval $\big(\frac{\alpha}{2\beta}, \widetilde{y}\big]$ since, otherwise, $F'' \geq 0$ on $\big[\widetilde{y}, K\big]$. By the Mean Value Theorem we know that there exists $\xi \in (\widetilde{y}, y)$ such that

$$F'(y) = F'(\widetilde{y}) + F''(\xi)(y - \widetilde{y}) = F''(\xi)(y - \widetilde{y}) \geq 0;$$

a contradiction. By putting all together we see that there exists a point $t \in (\widetilde{y}, K)$ such that $F''(t) = 0$, $F'' < 0$ on the interval $\big(\frac{\alpha}{2\beta}, t\big)$ and $F'' > 0$ on the interval (t, K). Again by the Mean Value Theorem we have that $F' < 0$ on $(\widetilde{y}, t]$. That is, there exists a unique intersection point of the curves $-f'(x) = 2\beta x - \alpha$ and $-\lambda\frac{d}{dx}D(x, \mu, \sigma, \delta)$ in the interval $\big(\frac{\alpha}{2\beta}, t\big]$. Analogously, $F'(K) \geq 0$, and there exists an intersection point of the curves in the interval $(t, K]$. Since F'' is positive in this interval this point must be unique.

(C) When $\delta \in \big(\frac{\alpha}{2\beta}, K\big)$ and $\mu < 2 + \frac{\sigma\delta}{\Theta}$ we have $\kappa \in \{0, 1, 2, 3\}$.
This follows by using (A) on the interval $[\delta, K]$ where the right hand side of Equation (4.9) is strictly decreasing, and by using (B) on $\big[\frac{\alpha}{2\beta}, \delta\big]$ where it is strictly increasing.

Then the proof follows by noticing that (A) fits into Statements (a,b), (B) fits into Statements (a,b,c,d), and (C) fits into Statements (a,b,c,d,e) because when $\kappa = 3$ none of the critical points can be an inflexion point. □

From the above two lemmas we get:

Corollary 4.6. *The vector field $F\big|_{[0,K]}$ has at most 4 stationary states (zeroes) and every possible cardinality of stationary states can be realized with non-degenerate zeroes (that is, zeroes where the map F is locally strictly monotone) with appropriate parameter values. Consequently the potential function of the vector field F has at most 4 critical points (of course, at most two maxima and at most two minima).*

4.2 Model fitting and parameters estimation: Collapse phase 2006–2017

Here, we consider the period from 2006 to 2017, thus focusing on the collapse phase involving the dispersal of almost all the individuals present at the patch of study. In what follows we will consider Model (4.1) with the parameters computed in Section 3.2.2, since these estimations from the initial phase are considered as the intrinsic population's characteristics. The solution of Model (4.1) with β fixed to $2.4382635446 \times 10^{-5}$, and its parameters belonging to the ranges shown in the table at the beginning of Section 4 (Page 15) will be denoted by $x(t) = x_{\varphi,\lambda,\mu,\sigma,\delta}(t)$, $t \in [0, 11]$. Observe that the solution $x(t)$ depends on the initial condition $x(0) = x_{\varphi,\lambda,\mu,\sigma,\delta}(0) \in [0, K]$, that must be considered as a free parameter as well.

On the other hand, we denote the observed population of Audouin's gulls at the years 2006 to 2017 by

$$\pi(\ell, \ell = 0:11) = \mathsf{Audouin's\_Gulls\_Observed\_Population\_at\_year}(2006 + \ell, \ell = 0:11) =$$
$$\big[\overset{[0]}{15329}, \overset{[1]}{14177}, \overset{[2]}{13031}, \overset{[3]}{9762}, \overset{[4]}{11271}, \overset{[5]}{8688}, \overset{[6]}{7571}, \overset{[7]}{6983}, \overset{[8]}{4778}, \overset{[9]}{2067}, \overset{[10]}{1586}, \overset{[11]}{793}\big].$$

Now we define the parameter space $\mathscr{P} := [0, K] \times [-\infty, \alpha] \times \mathbb{R}^+ \times \mathbb{R}^+ \times \mathbb{R}^+ \times \mathbb{R}^+$, and a map

$$\begin{array}{rccc} \mathsf{F}\colon & \mathscr{P} & \longrightarrow & \mathbb{R}^+ \\ & \big(x(0), \varphi, \lambda, \mu, \sigma, \delta\big) & \longmapsto & \sqrt{\sum_{\ell=0}^{11} \big(x(\ell) - \pi(\ell)\big)^2}. \end{array}$$

The *fitting of the model* consists in solving

$$\begin{aligned} &\min \mathsf{F}\big(x(0), \varphi, \lambda, \mu, \sigma, \delta\big) \\ &\text{subject to } \big(x(0), \varphi, \lambda, \mu, \sigma, \delta\big) \in \mathscr{P}, \\ &\text{and } x(t) \in [0, K] \text{ for } t \in [0, 11], \end{aligned} \tag{4.12}$$

and checking that this minimum is as low as possible to guarantee the validity of the model. As it has been noted in the case of the initial phase data in Section S3.2, the set

$$\Big\{\mathsf{F}\big(x(0), \varphi, \lambda, \mu, \sigma, \delta\big) : \big(x(0), \varphi, \lambda, \mu, \sigma, \delta\big) \in \mathscr{P} \text{ and } x(t) \in [0, K] \text{ for } t \in [0, 11]\Big\} \subset \mathbb{R}^+$$

has 0 as a lower bound. Hence, it has a minimum element, and Problem (4.12) has at least one solution.

Remark 4.7. The evaluation of the function $\mathsf{F}\big(x(0), \varphi, \lambda, \mu, \sigma, \delta\big)$ has to go through the computation of a solution of the ODE (4.1) whose dispersal term is highly nonlinear. When $\lambda > 0$ the approximate solution of (4.1) is computed numerically using the Runge-Kutta-Fehlberg-Simó integrator of order 7–8 with adaptive step-size for speed and efficiency. The version of the integrator that we use has been specially implemented for more optimal speed by one of the authors.

4.2.1 A first approach to fit the collapse phase: Montecarlo and Sparse Anisotropic Grid Searches

As said above, as a first approach to find the solutions of Problem (4.12), it is convenient to perform a brute force exploration of a reasonable region of the parameter space $\mathscr{P}$. The motivation for this exploration is twofold: first, to have a rough idea of the landscape (graph) of the function F, and, second, to find a point in $\mathscr{P}$, reasonably close to the optimum of Problem (4.12). This point will be used as a fulcrum to determine a compact relatively small set $\mathscr{K} \subset \mathscr{P}$ that contains the minimum (or equivalently that $\mathscr{P} \setminus \mathscr{K}$ does not contain the minimum) of function F. The existence of the compact set $\mathscr{K}$ has two important consequences. First, Bolzano-Weierstrass Theorem tells us that the fitness function F has a minimum in $\mathscr{K}$. Thus, by the choice of $\mathscr{K}$, this minimum must be the solution of Problem (4.12). Second, the reduction of the parameters' search space from $\mathscr{P}$ to $\mathscr{K}$ will make possible the minimization algorithms.

The *Grid Searching Method* has been implemented (after several numerical experiments) sparse and anisotropic on a reasonably small compact subregion of $\mathscr{P}$ with a relatively small computational complexity (i.e. the number of evaluations of the function F). The need for the compacity of the search domain is obvious. The reduction of the computational complexity of the grid search is clearly achieved by choosing a sparse grid but also by the anisotropy. By *anisotropy* we mean that, for certain parameters, the step used to construct the grid is not constant. It rather depends on the zone where the parameter lies, and on the desired precision in that zone.

At a first step, the ranges of parameters that determine the compact domain and their sparseness and anisotropy have been chosen arbitrarily (after several preliminary explorations with low computational complexity) since we only want to have a rough idea of the landscape (graph) of the function F and to find a point from $\mathscr{P}$ reasonably close to the optimum.

The *Sparse Anisotropic Grid Search (SAGS)* is completely specified (including the compact subregion of $\mathscr{P}$ where it is performed) in the left half of Table 2 below, and the obtained results are summarized in the next lemma.

Lemma 4.8. *We have*

$$\mathsf{F}\big(15800, 0.22, 1400, 0, 1, 8740\big) = 2602.4358676183260\cdots$$

and $x_{(15800,0.22,1400,0,1,8740)}(t) \in [0, K]$ *for* $t \in [0, 11]$.

In view of the above lemma, it makes sense to graphically explore the vicinity of the optimum point found by the anisotropic grid search to measure the "landscape complexity" of this vicinity. This is done in Figure S14 below.

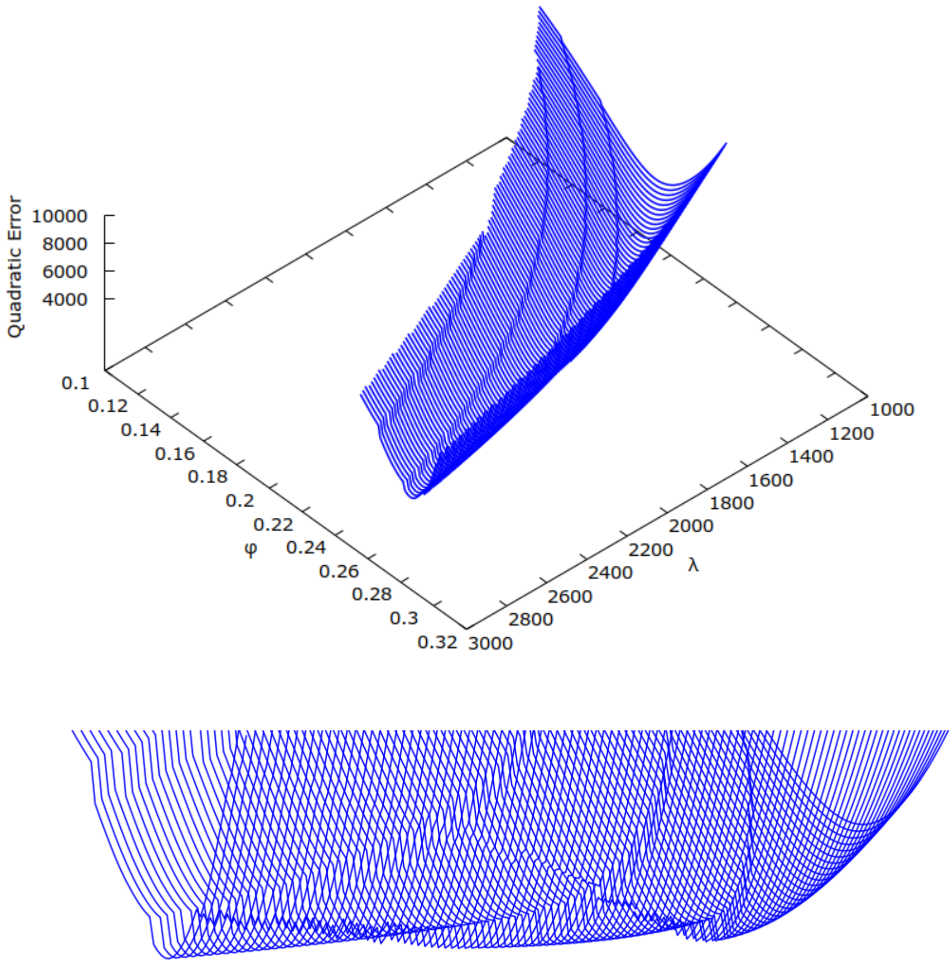


Figure S14. A view of the landscape of the function F around the point $(15800, 0.22, 1400, 0, 1, 8740)$ showing only points whose $\mathsf{F}-$value is lower than 10000. In the plots we have fixed the following 4 parameters: $x(0) = 15800$, $\mu = 0$, $\sigma = 1$ and $\delta = 8740$, λ ranges from 1000 to 3000 while φ ranges from 0.12 to α. The lower plot shows a zoom of the "valley" of the landscape from the upper picture where a complicate local minima distribution is seen.

A first attempt to compute the solution of Problem (4.12) has been to improve the SAG search by performing a Nonlinear Least-Squares Fitting by using a Levenberg-Marquardt Trust-Region Algorithm taking as seed the SAGS optima shown in Lemma 4.8. Unsurprisingly (see Figure S14) this latter algorithm has not been able to improve the previously found SAGS optima (even it has not been able to improve perturbed versions of the seed, used to increase the exploratory character of the whole search).

Conclusion: *In this search/computation we greatly need to increase the exploratoryness of our algorithms. This leads us in a natural way to Genetic Algorithms tuned to be highly exploratory.* The Genetic Algorithm that we will use in the Sub-subsection 4.2.3 and its computational efficiency will greatly benefit from the reduction of the search domain to a small compact set $\mathscr{K}$.

The determination of the above mentioned compact set, and the arguments to justify that $\mathscr{P} \setminus \mathscr{K}$ does not contain the minimum of Problem (4.12) will take advantage of a finer stratification of the domain $\mathscr{P} \setminus \mathscr{K}$. To this end we define

$$\begin{aligned}
\mathscr{S} &:= [12726, 17932] \times [0.12, \alpha] \times [300, +\infty] \times \mathbb{R}^+ \times \mathbb{R}^+ \times \mathbb{R}^+ \subset \mathscr{P} \\
\mathscr{L} &:= [12726, 17932] \times [0.12, \alpha] \times [300, 100000] \times [0, 600] \times [0, 5000] \times [0, 159000] \subset \mathscr{S}, \text{ and} \\
\mathscr{K} &:= [12726, 17932] \times [0.12, \alpha] \times [300, 3000] \times [0, 10] \times [0, 50] \times [0, 20000] \subset \mathscr{L}.
\end{aligned}$$

In this framework we have performed a *Sparse Anisotropic Large Domain Grid Search (SALDGS)* on the domain $\mathscr{L} \setminus \mathscr{K}$, completely specified in the right half of Table 2 above, and the obtained results are summarized in the next lemma.

Lemma 4.9. *For every point $\vec{\boldsymbol{\theta}} \in \mathscr{L} \setminus \mathscr{K}$ whose components belong to the grid described in Table 2 we have*

$$\mathsf{F}\big(\vec{\boldsymbol{\theta}}\big) > 2664 > \mathsf{F}\big(15800, 0.22, 1400, 0, 1, 8740\big).$$

Parameter	Theoretical Range	Sparse Anisotropic Grid Search: Effective Search Range	Sparse Anisotropic Grid Search: Anisotropic Step	Large Domain Grid Search (SALDGS): Effective Search Range	Large Domain Grid Search (SALDGS): Anisotropic Step
$x_{\varphi,\nu,\lambda,\mu,\sigma,\delta}(0)$	$[0, K]$	$[12600, 18800]$	200	$\mathscr{L} \setminus \mathscr{K}$	100
φ	$(-\infty, \alpha]$	$[0.13, 0.34]$	0.01		0.05
λ	$\mathbb{R}^+$	$[300, 3000]$	100		$\begin{cases} 100 & \text{when } \lambda \in [300, 2900], \\ 1000 & \text{when } \lambda \in [3000, 100000] \end{cases}$
μ	$\mathbb{R}^+$	$[0, 10]$	$\begin{cases} 0.1 & \text{when } \mu \in [0, 0.9], \\ 1 & \text{when } \mu \in [1, 10] \end{cases}$		$\begin{cases} 0.1 & \text{when } \mu \in [0, 0.9], \\ 1 & \text{when } \mu \in [1, 49], \\ 10 & \text{when } \mu \in [50, 90], \\ 100 & \text{when } \mu \in [100, 600], \end{cases}$
σ	$\mathbb{R}^+$	$[0, 50]$	$\begin{cases} 1 & \text{when } \sigma \in [0, 10], \\ 5 & \text{when } \sigma \in [11, 50] \end{cases}$		$\begin{cases} 0.1 & \text{when } \sigma \in [0, 1.9], \\ 2 & \text{when } \sigma \in [2, 38], \\ 10 & \text{when } \sigma \in [40, 90], \\ 100 & \text{when } \sigma \in [100, 5000] \end{cases}$
δ	$\mathbb{R}^+$	$[0, 20000]$	10		1000

Table 2. Left half: Full specification of the Sparse Anisotropic Grid Search (SAGS). For every parameter it is given the *effective search range* together with the step (anisotropic in the case of μ and σ) used in the search. The SAGS has explored 14,988,610,560 mesh points or, equivalently, it has evaluated the function $\mathsf{F}\big(x(0), \varphi, \lambda, \mu, \sigma, \delta\big)$ at 14,988,610,560 points of the feasible space $\mathscr{P}$.
Right half: Full specification of the Sparse Anisotropic Large Domain Grid Search (SALDGS). As for the SAGS case, for every parameter it is given the *effective search range* together with the step (anisotropic in the case of λ, μ and σ) used in the search. The SALDGS has explored 34,004,017,950 mesh points in the domain $\mathscr{L}$.

On the other hand, we also have performed a Montecarlo exploration on the computer-representable part of the region $\mathscr{S} \setminus \mathscr{L}$, to get

$$\mathsf{F}\big(\vec{\boldsymbol{\theta}}\big) > \mathsf{F}\big(15800, 0.22, 1400, 0, 1, 8740\big)$$

for every selected point $\vec{\boldsymbol{\theta}} \in \mathscr{S} \setminus \mathscr{L}$.

4.2.2 Analytic and heuristic estimates of a compact domain that contains the optimum

The goal of this subsection is to justify that the search space for solving Problem (4.12) can be greatly reduced to the compact set $\mathscr{K}$.

To this end, we introduce the *reduced minimization problem*

$$\begin{aligned} &\min \mathsf{F}\big(x(0), \varphi, \lambda, \mu, \sigma, \delta\big) \\ &\text{subject to } \big(x(0), \varphi, \lambda, \mu, \sigma, \delta\big) \in \mathscr{K}, \\ &\text{and } x(t) \in [0, K] \text{ for } t \in [0, 11], \end{aligned} \tag{4.13}$$

and we will semi-analytically justify (with the help of an heuristic reasoning) that the next results holds:

Proposition 4.10. *The solutions of Problem* (4.12) *and Problem* (4.13) *coincide.*

To justify Proposition 4.10 we will use the following technical lemma.

Lemma 4.11. *Let $\vec{\boldsymbol{\theta}} \in \mathscr{P} \setminus \mathscr{S}$. Then,*

$$\mathsf{F}\big(\vec{\boldsymbol{\theta}}\big) > \mathsf{F}\big(15800, 0.22, 1400, 0, 1, 8740\big).$$

Consequently,

$$\underset{\vec{\boldsymbol{\theta}} \in \mathscr{P}}{\arg\min} \, \mathsf{F}\big(\vec{\boldsymbol{\theta}}\big) \in \mathscr{S}.$$

Remark 4.12. The above lemma shows that the solution of Problem 4.12 verifies $\lambda \geq 300 \gg 0$, thus proving analytically the hypothesized highly nonlinear migratory behaviour of both, the Audouin's gulls and the theoretical model.

Justification of Proposition 4.10. From Lemma 4.11 we see that

$$\min\Big\{\mathsf{F}\big(\vec{\boldsymbol{\theta}}\big) \,:\, \vec{\boldsymbol{\theta}} \in \mathscr{P} \text{ and } x_{\vec{\boldsymbol{\theta}}}(t) \in [0, K] \text{ for } t \in [0, 11]\Big\} = \\ \min\Big\{\mathsf{F}\big(\vec{\boldsymbol{\theta}}\big) \,:\, \vec{\boldsymbol{\theta}} \in \mathscr{S} \text{ and } x_{\vec{\boldsymbol{\theta}}}(t) \in [0, K] \text{ for } t \in [0, 11]\Big\}.$$

The results of the Montecarlo exploration of the region $\mathscr{S} \setminus \mathscr{L}$, Lemma 4.9, and a continuity argument heuristically give

$$\min\Big\{\mathsf{F}\big(\vec{\boldsymbol{\theta}}\big) \,:\, \vec{\boldsymbol{\theta}} \in \mathscr{S} \text{ and } x_{\vec{\boldsymbol{\theta}}}(t) \in [0, K] \text{ for } t \in [0, 11]\Big\} = \\ \min\Big\{\mathsf{F}\big(\vec{\boldsymbol{\theta}}\big) \,:\, \vec{\boldsymbol{\theta}} \in \mathscr{L} \text{ and } x_{\vec{\boldsymbol{\theta}}}(t) \in [0, K] \text{ for } t \in [0, 11]\Big\} = \\ \min\Big\{\mathsf{F}\big(\vec{\boldsymbol{\theta}}\big) \,:\, \vec{\boldsymbol{\theta}} \in \mathscr{K} \text{ and } x_{\vec{\boldsymbol{\theta}}}(t) \in [0, K] \text{ for } t \in [0, 11]\Big\}.$$

This ends the justification of Proposition 4.10. $\square$

To prove Lemma 4.11 we will use the following analytical result.

Lemma 4.13. *Let $f, g \colon \mathbb{R}^+ \longrightarrow \mathbb{R}$ be continuous functions, and let $x(t)$ and $y(t)$ denote the solutions of the differential equations $\frac{\mathrm{d}}{\mathrm{d}t}\, x(t) = f\big(x(t)\big)$ and $\frac{\mathrm{d}}{\mathrm{d}t}\, y(t) = g\big(y(t)\big)$ with initial conditions $x(0)$ and $y(0)$, respectively. Assume that the solution $x(t)$ is defined and non-negative (i.e. $x(t) \in \mathbb{R}^+$) for every t in an interval $[0, T]$.*

(a) Suppose that, $x(0) \leq y(0)$ and $f(x) \leq g(x)$ for every $x \in \mathbb{R}^+$. Then, $y(t)$ is defined for every t in the interval $[0, T]$, and $x(t) \leq y(t)$ for every $t \in [0, T]$.

(b) Suppose that, $0 \leq y(0) \leq x(0)$ and $g(x) \leq f(x)$ for every $x \in \big[0, \max_{t \in [0,T]} x(t)\big]$. Then there exists a maximal interval $[0, T^*] \subset [0, T]$ such that $y(t)$ is defined for every $t \in [0, T^*]$, and $0 \leq y(t) \leq x(t)$ for every $t \in [0, T^*]$.

The proof of this lemma uses heavily the *Fundamental Theorem of Calculus* which, in this case, gives

$$x(t) = x(t_0) + \int_{t_0}^{t} f(x(s))\,\mathrm{d}s \tag{4.14}$$

whenever the solution $x(t)$ exists and is bounded and non-negative in an interval $[t_0, t] \subset [0, T]$.

Remark 4.14. A sufficient condition for $T^* < T$ in Lemma 4.13(b) is that $y(T^*) = 0$ and $g(0) < 0$.

Now we are ready for the

Proof of Lemma 4.11. Let $\vec{\boldsymbol{\theta}} = (\kappa, \varphi, \lambda, \mu, \sigma, \delta) \in \mathscr{P} \setminus \mathscr{S}$. We start by assuming that $\kappa \leq 12726$. By Lemma 4.8 we have,

$$\mathsf{F}\big(\kappa, \varphi, \nu, \lambda, \mu, \sigma, \delta\big) = \sqrt{\sum_{\ell=0}^{11} \big(x(\ell) - \pi(\ell)\big)^2} \geq \sqrt{\big(\pi(0) - x(0)\big)^2} =$$

$$\pi(0) - \kappa \geq 15329 - 12726 > \mathsf{F}\big(15800, 0.22, 1400, 0, 1, 8740\big).$$

Analogously, if $\kappa \geq 17932$,

$$\mathsf{F}\big(\kappa, \varphi, \nu, \lambda, \mu, \sigma, \delta\big) \geq \kappa - \pi(0) \geq 17932 - 15329 > \mathsf{F}\big(15800, 0.22, 1400, 0, 1, 8740\big).$$

Thus, in what follows we may assume that $x(0) = \kappa \in (12726, 17932)$.

Now assume that $\varphi \leq 0.12$. We denote by $u(t)$, $t \in [0, 11]$, the solution of Model (3.1) with α replaced by $\widetilde{\varphi} = 0.12$, $\beta = 2.43826356697 \times 10^{-5}$, and initial condition $u(0) = 17932$. By Lemma 3.1,

$$u(t) = \frac{\widetilde{\varphi} u(0) \exp(\widetilde{\varphi} t)}{\widetilde{\varphi} + \beta u(0)(\exp(\widetilde{\varphi} t) - 1)},$$

which is clearly defined, non-negative and bounded on the interval $[0, 11]$. By direct computation, we get

$$\begin{array}{lcl}
u(1) \approx 13805.1588980\cdots & < & \pi(1) = 14177, \\
u(2) \approx 11464.9892996\cdots & < & \pi(2) = 13031, \\
u(4) \approx 8931.2750\cdots & < & \pi(4) = 11271, \\
u(5) \approx 8177.851882950\cdots & < & \pi(5) = \ \ 8688,
\end{array}$$

and

$$\sqrt{\sum_{\ell \in \{1,2,4,5\}} \big(\pi(\ell) - u(\ell)\big)^2} = 2885.34\cdots\ .$$

By Proposition 4.1(a),

$$\varphi x - \beta x^2 - \lambda D(x, \mu, \sigma, \delta) \leq \widetilde{\varphi} x - \beta x^2,$$

for every $x \in \mathbb{R}^+$. Then, since $x(0) \leq 17932 = u(0)$, by Lemma 4.13(b) we get that either $x(t)$ is not defined for every t in the interval $[0, 11]$ (in particular $x(t)$ is not feasible), or

$$x(\ell) \leq u(\ell) < \pi(\ell)$$

for $\ell = 1, 2, 4, 5$. Hence, by Lemma 4.8,

$$\mathsf{F}\big(\kappa, \varphi, \nu, \lambda, \mu, \sigma, \delta\big) = \sqrt{\sum_{\ell=0}^{11} \big(x(\ell) - \pi(\ell)\big)^2} \geq \sqrt{\sum_{\ell \in \{1,2,4,5\}} \big(\pi(\ell) - x(\ell)\big)^2} \geq$$

$$\sqrt{\sum_{\ell \in \{1,2,4,5\}} \big(\pi(\ell) - u(\ell)\big)^2} > 2885 > \mathsf{F}\big(15800, 0.22, 1400, 0, 1, 8740\big).$$

So, in what follows we additionally may assume that $\varphi > 0.12$.

Next we denote by $v(t)$, $t \in [0, 11]$, the solution of

$$\frac{dv(t)}{dt} = \widetilde{\varphi} v(t) - \beta v(t)^2 - \nu, \tag{4.15}$$

with $\widetilde{\varphi} = 0.12$, $\beta = 2.43826356697 \times 10^{-5}$, $\nu = 600$ and initial condition $v(0) = 12726$. By direct computation, we get

$$\begin{array}{lll} v(9) \approx 3461.6330\cdots & > & \pi(9) = 2067, \\ v(10) \approx 2994.53466770\cdots & > & \pi(10) = 1586, \\ v(11) \approx 2539.4820\cdots & > & \pi(11) = 793, \end{array}$$

(in particular $v(t)$ is defined, non-negative and bounded on the interval $[0, 11]$), and

$$\sqrt{\sum_{\ell \in \{9,10,11\}} \big(\pi(\ell) - v(\ell)\big)^2} = 2641.8120\cdots .$$

By Proposition 4.1 we have

$$D(x, \mu, \sigma, \delta) \leq \begin{cases} D(0, \mu, \sigma, \delta) = 1 < 2 & \text{when } \mu \geq 1, \text{ and} \\ D(x^*, \mu, \sigma, \delta) < 2 & \text{when } 0 \leq \mu < 1. \end{cases}$$

Consequently, when $\lambda \leq 300$ we have $\lambda D(x, \mu, \sigma, \delta) < 2\lambda \leq \nu$, and

$$\varphi x - \beta x^2 - \lambda D(x, \mu, \sigma, \delta) > \widetilde{\varphi} x - \beta x^2 - \nu$$

for every $x \in \mathbb{R}^+$. Then, since $x(0) > 12726 = v(0)$, by Lemma 4.13(a),

$$x(\ell) \geq v(\ell) > \pi(\ell)$$

for $\ell = 9, 10, 11$. Hence, by Lemma 4.8,

$$\begin{aligned} \mathsf{F}\big(\kappa, \varphi, \nu, \lambda, \mu, \sigma, \delta\big) = \sqrt{\sum_{\ell=0}^{11} \big(x(\ell) - \pi(\ell)\big)^2} \geq \sqrt{\sum_{\ell \in \{9,10,11\}} \big(\pi(\ell) - x(\ell)\big)^2} \geq \\ \sqrt{\sum_{\ell \in \{9,10,11\}} \big(\pi(\ell) - v(\ell)\big)^2} > 2641 > \mathsf{F}\big(15800, 0.22, 1400, 0, 1, 8740\big). \end{aligned}$$

□

4.2.3 Fitting the collapse phase using artificial intelligence: Genetic Algorithms

As it has been already explained, we want to minimize the function F (or, ideally, to find a vector of parameters $\vec{\boldsymbol{\theta}} \in \mathscr{P}$ such that $\mathsf{F}(\vec{\boldsymbol{\theta}}) = 0$). This amounts solving Problem 4.12 which, in view of Proposition 4.10, is equivalent to solve the *reduced minimization problem*

$$\begin{aligned} &\min\ \mathsf{F}\big(x(0), \varphi, \lambda, \mu, \sigma, \delta\big) \\ &\text{subject to } \big(x(0), \varphi, \lambda, \mu, \sigma, \delta\big) \in \mathscr{K}, \\ &\text{and } x(t) \in [0, K] \text{ for } t \in [0, 11]. \end{aligned} \tag{4.2}$$

To solve Problem (4.13) we will use the following Standard Genetic Algorithm (GA) tuned to be highly exploratory, with F as its fitness function. A diagram of the GA displayed in Box 1 below with the pseudo-code, the used variables, the implemented functions, and the execution flow.

Box 1. A Standard Exploratory GA

```
popsize ← desired population size                                   ▷ Must be even
P ← {}                                                  ▷ Initializing empty first generation
for popsize times do
    P ← P ∪ {new random individual}
end for                                                 ▷ End of first generation initialization

procedure GeneticAlgorithm(P, popsize)
    bestfitness ← ∞                                                 ▷ bestfitness initialization
    repeat
        for each individual p ∈ P do
            fit ← F(p)
            if fit < bestfitness then        ▷ True when bestfitness is ∞. Found best individual so far
                bestfitness ← fit
                Best ← p
            end if
        end for
        Q ← {}                                          ▷ Initializing empty next generation
        for popsize/2 times do
            Parent p_a ← SelectWithReplacement(P)
            Parent p_b ← SelectWithReplacement(P)
            Children c_a, c_b ← Crossover(Copy(p_a), Copy(p_b))
            Q ← Q ∪ {Mutate(c_a),Mutate(c_b)}
        end for
        P ← Q           ▷ The population is replaced completely at each generation: we want to be exploratory
    until Best is the ideal solution or the maximum number of generations has been exhausted
    return Best
end procedure
```

To completely specify such an algorithm, a number of its elements have to be defined and revised. Namely: how individuals are coded, the operations SelectWithReplacement, Crossover, Mutate, and finally the stopping criteria. Also, a number of parameters of the algorithm have to be introduced and discussed: the population size (*popsize*), the maximum number of generations and the mutation probability. Additionally, the stopping criteria and the function select with replacement (SelectWithReplacement) depend on internal parameters that will be explained whenever these two procedures are surveyed.

The next subsections will be devoted to explain our implementation of the items listed above.

Population parameters and stopping criteria

Since the algorithm must be tuned to be highly exploratory, it is necessary to have large *popsize* and number of generations. We have taken 20000 as tentative value for *popsize* (although this will be better dealt in our variant of the algorithm developed in Section 4.2.3), and the maximum number of generations *MaxNumGen* is taken to be 1000. Accordingly, the mutation probability has been set to 0.1 (it must be small but not too small for exploratoriness).

The basic stopping criteria depends on a new parameter which controls the maximum number of generations without improving *bestfitness*. This parameter, called, *MaxNumGenWithoutImproving* is set to 100. It works essentially as indicated by its name: after 100 generations without improvement (or modification) of *bestfitness* the algorithm stops and returns the best individual found. This can be taken as our definition of the term ideal solution in the above version of the GA.

Individuals

An individual in the population is specified by six "genes" corresponding to the six free parameters: the initial condition $x(0)$, φ, λ, μ, σ and δ. In the framework established by the proof of Holland's Convergence Theorem [17] it is convenient to write the genes as unsigned integers expressed in binary, with its range depending on the true range of the real parameters and its sensitivity. This leads to the distinction between the individual's *phenotype* which corresponds to the real (human readable) parameter values, and the *(discretized) genotype* which is composed of the same parameters but written as unsigned integers in binary. Of course the *translation procedures* from phenotype to genotype and vice-versa must be specified. In Table 3 we specify, for each phenotypic parameter, its *Theoretical Range*, the associated *Effective Search Range* determined by the set $\mathscr{S}$ (which is the feasible space of Problem (4.13)), a reasonable sensitivity, or, better said, *Precision or Discretization Step* (which determines the precision of the parameter's estimate), and the associated genotype which consists on two elements: the *Unsigned Integer Upper Limit* which, as we will see, should be always taken as a power of two 2^ϱ (where ϱ depends on the effective search range and discretization step — it determines also an *unsigned integer search range* of the form $0, 1, 2, \dots, 2^\varrho - 1$), and the *translation function* from genotype to phenotype.

Table 3. Full specification of individuals for the GA and their genes coding

		Phenotype		Genotype		
Parameter	Theoretical Range	Effective Search Range	Precision or Discretization Step	Unsigned Integer Upper Limit	Effective Discretization Step	Translation map from genotype to phenotype
$x(0)$	$[0, K]$	$[12726, 17932]$	10^{-2}	2^{19}	$\frac{5206}{2^{19}-1}$	$u \longmapsto 12726 + u\ \frac{5206}{2^{19}-1}$
φ	$(-\infty, \alpha]$	$[0.12, \alpha]$	10^{-10}	2^{32}	$\frac{\alpha-0.12}{2^{32}-1}$	$u \longmapsto 0.12 + u\ \frac{\alpha-0.12}{2^{32}-1}$
λ	$\mathbb{R}^+$	$[300, 3000]$	10^{-2}	2^{19}	$\frac{2700}{2^{19}-1}$	$u \longmapsto 300 + u\ \frac{2700}{2^{19}-1}$
μ	$\mathbb{R}^+$	$[0, 10]$	10^{-6}	2^{24}	$\frac{10}{2^{24}-1}$	$u \longmapsto u\ \frac{10}{2^{24}-1}$
σ	$\mathbb{R}^+$	$[0, 50]$	10^{-4}	2^{19}	$\frac{50}{2^{19}-1}$	$u \longmapsto u\ \frac{50}{2^{19}-1}$
δ	$\mathbb{R}^+$	$[0, 20000]$	0.5	2^{16}	$\frac{20000}{2^{16}-1}$	$u \longmapsto u\ \frac{20000}{2^{16}-1}$

Remark 4.15 (On the determination of the Unsigned Integer Upper Limit and the Effective Discretization Step)**.** Assume that a parameter has an *Effective Search Range* of the form $[A, B]$ and a desired *Precision or Discretization Step* ξ. The integer range corresponding to A, B, and ξ is $0, 1, 2, \dots, \left\lceil \frac{B-A}{\xi} \right\rceil$, where $\lceil\cdot\rceil$ denotes the *ceiling function*. The *Genotype Unsigned Integer Upper Limit* is defined to be the smallest power of two 2^ϱ such that $\left\lceil \frac{B-A}{\xi} \right\rceil \le 2^\varrho$.

Then, the available range of genotypic values for the parameter is $0, 1, 2, \dots, 2^\varrho - 1$, and the *Effective precision or Discretization Step* is $\frac{B-A}{2^\varrho-1}$. Consequently, the translation formula from genotype to phenotype is

$$u \longmapsto A + u\ \frac{B-A}{2^\varrho - 1},$$

and hence $0 \longmapsto A$ and $2^\varrho - 1 \longmapsto A + (2^\varrho - 1)\ \frac{B-A}{2^\varrho-1} = B$.

Remark 4.16. All *Genotype Unsigned Integer Upper Limits* in Table 3 above have exponent less than or equal to 32. This means that the base data type to store the genotypic values of all genes can be `unsigned int`'s of 32 bits.

Observe that *in this framework the restrictions on the parameters are verified automatically.* Indeed, we are restricting the genotypic values of a parameter to integers of the form $0, 1, 2, \dots, 2^\varrho - 1$ with a translation formula from genotype to phenotype of the form

$$u \longmapsto A + u\ \frac{B-A}{2^\varrho - 1} \in [A, B].$$

Since the *effective search ranges* in Table 3 are contained in the *Theoretical Ranges* and all values of parameters constructed by the GA are valid in the genotypic sense (i.e. belong to $\{0, 1, 2, \dots, 2^\varrho - 1\}$), the phenotypic parameter values must belong to the *Theoretical Ranges*, and hence verify all restrictions.

Remark 4.17 (On why we want the *Genotype Unsigned Integer Upper Limit* to be a power of two)**.** All genotypic values in the range $0, 1, 2, \dots, 2^\varrho - 1$, written in binary have a string of $32 - \varrho$ consecutive zeroes at the most significant bits part of the number, and a string of ϱ least significant bits. Eventually, for the number $2^\varrho - 1$, all ϱ least significant bits are set to 1. This eases the programming of crossovers and mutations, and will help avoiding complicate feasibility tests.

A final comment referring to the individuals' genotypes is that, for computational efficiency, is crucial to define an appropriate *data type* for them. In our case an individual is a `struct` composed by a vector of 6 *unsigned integers* (the genotype) and a variable to store the fitness value of the individual. This is accompanied (at the level of the whole population not of each individual) by the six translation formulae from genotype to phenotype shown in Table 3, and the list of exponents of the *Genotype Unsigned Integer Upper Limits* that, as explained in Remark 4.17, is crucial when setting the crossover and mutation procedures.

Finally, if we denote by $(u_{x(0)}, u_\varphi, u_\lambda, u_\mu, u_\sigma, u_\delta)$ the genes vector of an individual then, the *genotypic fitness function* is

$$(u_{x(0)}, u_\varphi, u_\lambda, u_\mu, u_\sigma, u_\delta) \longmapsto \mathsf{F}\Big(12726 + u_{x(0)}\,\tfrac{5206}{2^{19}-1}, 0.12 + u_\varphi\,\tfrac{\alpha-0.12}{2^{32}-1}, 300 + u_\lambda\,\tfrac{2700}{2^{19}-1}, u_\mu\,\tfrac{10}{2^{24}-1}, u_\sigma\,\tfrac{50}{2^{19}-1}, u_\delta\,\tfrac{20000}{2^{16}-1}\Big).$$

Selection with replacement

We use tournament algorithm with tournament parameter 10 (to increase exploratoriness). The detailed explanation of the procedure (in pseudocode) is the following:

Tournament Selection Algorithm

```
Require: P, t                      ▷ The population and the tournament size, t ≥ 1
  Best ← individual picked at random from P with replacement
  for i = 2 to t do
     p ← individual picked at random from P with replacement
     if F(p) < F(Best) then
        Best ← p
     end if
  end for
  return Best
```

Random initial population

Randomness in this setting is crucial although our variant of the algorithm will slightly improve — for good reasons — the initial population thus breaking its "pure randomness".

The initial population plays the role of a sample and, if it is not distributed uniformly in the whole search space, the optimum can be far from this initial sample and therefore missed[1] or, at least, the whole search can be delayed[2].

To assure the randomness of the initial population we avoid the use of congruential random number generators. We use a completely different approach. First we have designed a high quality random bits generator. This is done with a standard (i.e. not "high tech") random numbers generator modified for binary lotteries (i.e. giving only 0's and 1's). Then as a second step we use the random binary lotteries generator to perform lotteries in pairs and use the John von Neumann trick: if both results in the pair coincide, the roll is discarded; if, on the contrary, the results in the pair are different we take the first one as the generated resultant bit. This very clever von Neumann's strategy gives an unbiased random bits generator but it is somehow inefficient[3].

Equipped with the unbiased random bits generator, to build the initial population in generation zero, we fill the *Genotype Unsigned Integer Upper Limit exponent*–least significant bits of the six genes of every one of the *popsize* individuals.

Mutation

We use a very simple but aggressive mutation scheme (recall that we have to be highly exploratory). For *every gene (genotypic parameter) of every generated child* we swap a single random bit (among the *Genotype Unsigned Integer Upper Limit exponent*–least significant bits) with probability *MutationProbability* $= 0.1$.

[1] However, if the algorithm is exploratory enough, it can discover regions "hidden" to the initial population and "repare" this problem.

[2] An initial population well randomised is therefore a good investment in computational efficiency.

[3] The more biased it is the initial random binary lotteries generator, more rolls will have to be discarded and more inefficient the random bits generator becomes.

Crossover

Here, we perform one-point crossover among the *Genotype Unsigned Integer Upper Limit exponent*–least significant bits of *every gene (genotypic parameter)* of the two parents. This is best explained in the following picture

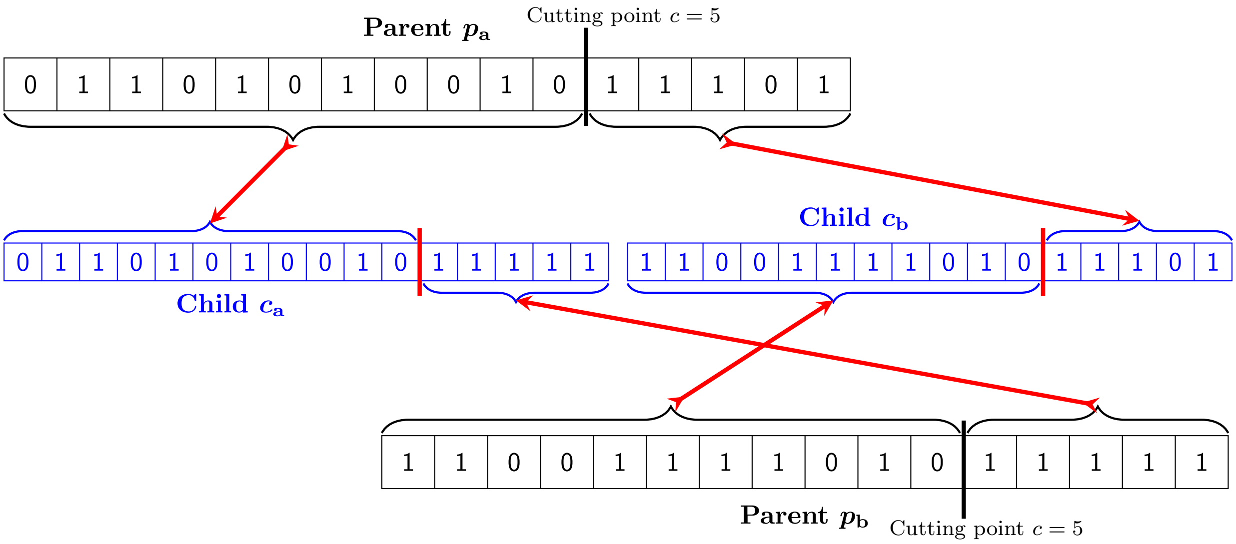


Figure S15. For both parents we show the *Genotype Unsigned Integer Upper Limit exponent*–least significant bits of the same gene (say gene 6 that would correspond to parameter δ) of the genotype. In accordance with Table 3, the *Genotype Unsigned Integer Upper Limit exponent* is 16. We also show the one-point crossover with cutting point at bit $c = 5$.

The *crossover cutting point* c is selected randomly for every gene among the least significant bits; but in a way that there is effective crossover (i.e. c must be different from 0 and the *Genotype Unsigned Integer Upper Limit exponent*). See Figure S15 above for a schematic diagram.

The Set of Superior variants of the Genetic Algorithm and its execution flow

Usually, the execution flow of a Genetic Algorithm (GA) is to run a batch of instances of the algorithm (in our case the *Standard Exploratory GA*) with different sets of algorithmic parameters (*popsize*, *MaxNumGen*, the mutation probability and others) and giving as a candidate to the optimum the best individual found in the whole batch.

However, as seen in Figure S14 (see also the *Full specification of individuals for the GA and their genes coding* table below) on the one hand, the search space is enormous: it has 2^{129} possible individuals and the dimesion three landscape given by φ, λ and F it has already a lot of very narrow local minima. We cannot imagine the complicacy of the landscape in dimension 7, taking into account the fact that some parameters (such as φ) are highly sensitive, while others, such as $x(0)$ and δ, have milder effects on the solution generated by the model.

These considerations tell us that finding a solution candidate to Problem (4.13) is really difficult and, as it has been said, it must be done with heuristic highly exploratory algorithms. However, as one can clearly see by looking at the results of the first batch of executions, it is good to "anchor" the now-not-so-random initial population to the fittest individuals when some is discovered. This adds an "elitist" ingredient to our algorithm for efficiency in local search. The fittest individuals just described are called the Superior ones, and the implementation of this idea gives a new meta-algorithm described in pseudo-code in the Box 2 below:

Box 2. GA in Recurring Batches with Initial Population Reinforced by the Set of Superiors

```
BestSuperior ← Fittest individual from Sparse Anisotropic Grid Search
BestSuperiorFitness ← F(BestSuperior)            ▷ Set of Superiors best fitness initialization
FF ← {BestSuperior}                              ▷ Set of Superiors initialization
while true do
    batchsize ← desired batch size for current iteration
    bestbatchfitness ← ∞                         ▷ Best batch fitness initialization
    for b ← 1 to batchsize do
        popsize_b ← desired population size for the b–batch iteration     ▷ Must be even
        P ← {}                                   ▷ Initializing empty first generation
        for popsize_b times do
            P ← P ∪ {new random individual}
        end for
        for each individual w ∈ FF do
            P ← w at a random place     ▷ Individuals from the Set of Superiors added at random places of P
        end for
        Best ← GeneticAlgorithm(P, popsize_b)    ▷ Standard Genetic Algorithm with P as initial population
        if F(Best) < bestbatchfitness then       ▷ Computing the best individual in the whole batch
            bestbatchfitness ← F(Best)
            BestInBatch ← Best
        end if
    end for                                      ▷ End of batch
    if bestbatchfitness < BestSuperiorFitness then   ▷ Updating the Set of Superiors, if necessary
        FF ← FF ∪ {BestInBatch}
        BestSuperior ← BestInBatch
        BestSuperiorFitness ← bestbatchfitness
    else
        return BestSuperior                      ▷ The End when there is no improvement
    end if
end while
```

Observe that in the above algorithm the Set of Superiors is formed by the fittest individuals of every batch and it is nitialized to the best phenotypic individual found in the Sparse Anisotropic Grid Search. Observe also that for every run of the *Standard Exploratory GA* in a batch, a single instance of every Superior individual is added to the now-not-so-random initial population at a random place (in particular a Superior individual can replace another Superior individual previously added to the initial population). In other words, the random initial population of every *Standard Exploratory GA* is anchored to the "optimal search space zone" by means of the Set of Superiors.

The results

In the table below we explain the execution flow of our *GAs in Recurring Batches with initial population reinforced by the Set of Superiors* which consists in the Sparse Anisotropic Grid Search and 6 batches.

The Set of Superiors of every batch is the result of the SAG search and a Best Batch Individual from every one of the previous batches (in the case of batch 4 we add two Superiors to the set instead of one because this batch gave a lot of better fitted individuals). Every row shows the best result from the batch (i.e. Best Batch Individual): columns 2–7 show the individual's phenotype and column 8 the individual's fitness (i.e. least-squares norm). The last batch (number 4) is used only as a stopping condition, i.e. to check that Best Batch Fitness does not improve.

Remark 4.18. The fact that in the last batch (i.e. for 4800 runs of the algorithm) we obtain a *unique Best Batch Individual*, and these coincide *for all population sizes and mutation probabilties*, tells us that *probably* the result we have found is the true global optimum of Problem (4.1).

In the next page we summarize the parameters and data corresponding to the optimum given in the last row of the table above (with blue background), including the data prediction from year 2006 to 2017 with a picture of the data fitting and the shape of the dispersal term

$$\lambda D(x, \mu, \sigma, \delta) = 1570.2313 \cdot D(x, 0.0, 0.49047, 8944.228).$$

First Superior: the output of the Sparse Anisotropic Grid Search (Lemma 4.8)							
Number of Fitness evaluations	$x(0)$	φ	λ	μ	σ	δ	Fitness
14,988,610,560	15800	0.22	1400	0	1	8740	2602.435867

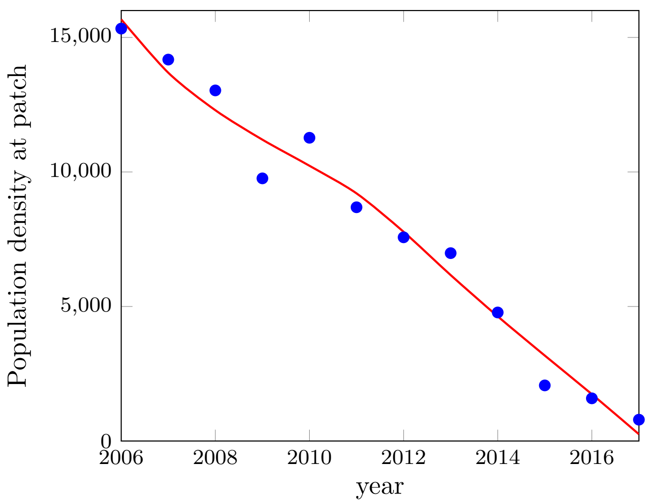
The rest of the Set of Superiors: a constant improvement of Best Batch Fitness							
#	$x(0)$	φ	λ	μ	σ	δ	Best Batch Fitness
1	15669.8212	0.2496035710078995	1569.8760	0.00001967	0.49219	8943.007	2567.00594448927
2	15670.5361	0.2497258272459134	1570.2416	0.0	0.49047	8944.228	2566.99966811847
3	15670.5560	0.2497248909716255	1570.2313	0.0	0.49047	8944.228	2566.99966764013
4	15670.5560	0.2497248909716255	1570.2313	0.0	0.49047	8944.228	2566.99966764013

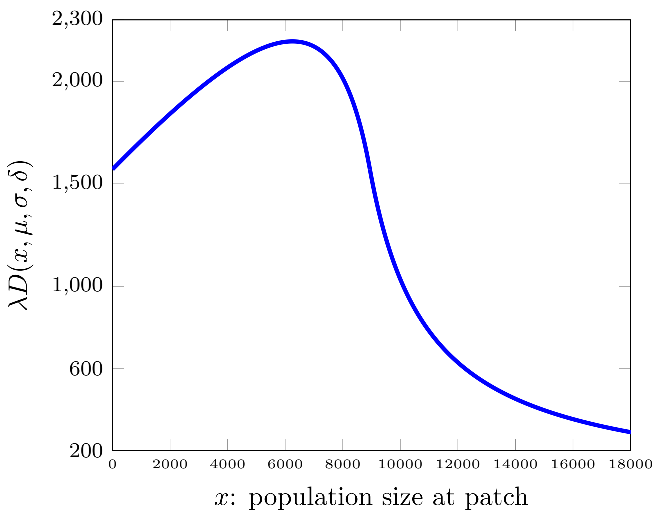
Table 4. A full account of the execution flow of the Genetic Algorithm (GA), together with the building of the final Set of Superiors. Every batch is divided into 8 sub-batches. In each sub-batch the population size is constant and ranges from 15000 to 29000 in steps of 2000. Every sub-batch consists on 600 executions of the GA with mutation probability 0.075, 0.1 and 0.15 (with 200 execusion for each mutation probability value). Thus, every batch has performed 4800 executions of the GA with different population sizes and mutation probabilities. Every row shows the result (i.e. Best Batch Individual) obtained in the corresponding batch: columns 2–7 show the individual's phenotype and column 8 the individual's fitness i.e., least-squares norm.

Remark 4.19 (The best fitting clearly does not correspond to density-independent dispersal)**.** Particularly, we observe that the parameters of the function $D(x, \mu, \sigma, \delta)$ corresponding to the best fitting fall away of the parameter region giving density-independent dispersal (see Remark 4.2). However, in Section 5 we shall check empirically this fact by showing that the best fitting for a density-independent dispersal is worst than the fitting obtained with social copying shown above.

Parameters' values at the optimum	
Parameter	**Value**
$x(0)$	15670.5560275192783593
φ	0.2497248909716255
β	$2.43826356697 \times 10^{-5}$
λ	1570.2313809039706030
μ	0.0
σ	0.4904756364357690
δ	8944.2282749675759987
Quadratic Error = $\mathsf{F}\big(x(0), \varphi, \lambda, \mu, \sigma, \delta\big)$	2566.999667640135158
Coefficient of determination R^2	$0.975715\cdots$

Year	Population Data	
	Observed	**Predicted**
2006	15329	15670.55
2007	14177	13688.02
2008	13031	12294.89
2009	9762	11200.30
2010	11271	10230.07
2011	8688	9203.86
2012	7571	7775.39
2013	6983	6167.47
2014	4778	4633.11
2015	2067	3176.01
2016	1586	1740.89
2017	793	252.86





4.3 A change in the tendency of gulls' population increase at the onset of perturbation

The aim of this section is to explore the change in the tendency of gulls' population increase coinciding with the onset of the perturbation, when predators arrived at the patch. This section has been placed after the analyses and computations carried out for the collapse phase because the fitting after the onset of the perturbation has been carried out using the parameters of the Elliot sigmoid function obtained in Section 4.2. Despite the amount of data for this period is very limited since the decline of the population ranges from 1998 to 2004, we will analyse this period of time considering dispersal. Before doing so, we will explore the period from the establishment of the population in 1981 to 2004, in order to see how a logistic model may provide or not a good fitting of the population dynamics for this period. Figure S16 displays the dynamics predicted by Equation (3.1) until 2004 using the structural population parameters estimated in Section 3.2.2. Notice that by extending the time series until 2004 using the estimated values of the initial phase, the field data after 2007 clearly deviates from the dynamics obtained with the parameter values before the arrival of predators.

Year	Population Data	
	Observed	**Predicted**
1997	11725	12096.688846
1998	11691	12674.456096
1999	10189	13116.386428
2000	10537	13447.142388
2001	11666	13690.683225
2002	10122	13867.859027
2003	10355	13995.627841
2004	9168	14087.185118

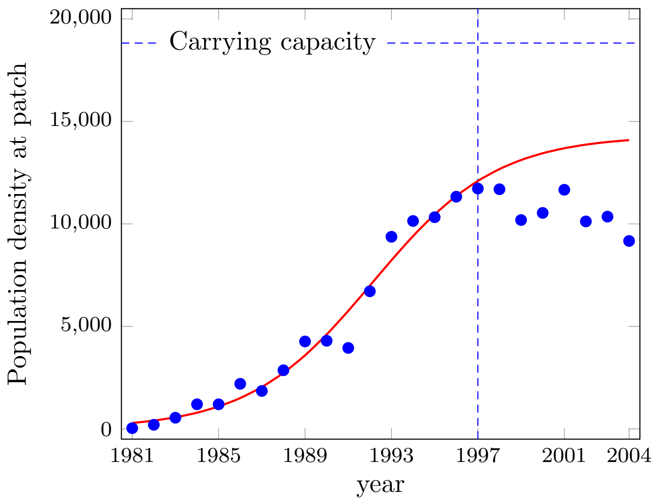


Figure S16. Left: Predictions for the period 1981–2004 using Equation (3.1) with the estimated structural parameters given in Section 3.2.2.
Right: Dynamics obtained with the structural parameters (red line). Field data shown with blue dots. Notice the change in the tendency at the onset of the perturbation.

Next, we fit the model to the period 1981–2004 using the same methodology than in Section 3.2.2. Here, we have used the value of β obtained in the initial phase ($\beta = 2.43826356697 \times 10^{-5}$), leaving as free parameters the initial condition and α. We have done this way to allow the population to decrease towards lower population values since $\alpha = \gamma - \varepsilon$. The best fit has been obtained for $x_0 = 603.140497$ and $\alpha = 0.2931419237955333$. However, the error is much higher than the one obtained for the initial phase ($\mathsf{LS}= 2593.053614$), now given by $\mathsf{LS}= 5639.340356$ for the period 1981–2004. The predicted versus the observed values are displayed in Figure S17.

Finally, we will fit the period 1998–2004 taking into account dispersal (both linear and by social copying). Here, as well, we do not aim at providing an exhaustive fitting for this short period of time, since the field data are scarce, but evaluate the tendency evaluating the weight of exponential dispersal versus dispersal by social copying. We want to emphasise that the period of interest is the local collapse observed from 2006 to 2017 investigated in Section S4.2.

To fit the dispersal by social copying we will use the parameters for the function $D(x, \mu, \sigma, \delta)$ obtained from the collapse period. We are doing so for two reasons. First, we are assuming that the shape of the dispersal function by social copying is an inherent trait of this species and thus the values of μ, σ and δ will be approximately constant. Second, we are only leaving λ as free parameter to avoid over-fitting, since the Elliot sigmoid function has three parameters. Figure S18 displays the best fit obtained for $\varphi = 0.2983089056790262$ and $\lambda = 1352.9233$. Also, the initial condition for this period must be $x(16) = x(1997) = 12096.688846$. More precisely, the model we consider for the period 1998–2004 is:

Observed & Predicted Population Data per Year								
Year	**Obs**	**Predicted**	**Year**	**Obs**	**Predicted**	**Year**	**Obs**	**Predicted**
1981	36	603.140497	1989	4266	4271.719874	1997	11725	10241.674836
1982	200	795.004520	1990	4300	5108.525129	1998	11691	10642.218528
1983	546	1042.331130	1991	3950	5982.725038	1999	10189	10962.004166
1984	1200	1357.299569	1992	6714	6858.134649	2000	10537	11213.338485
1985	1200	1752.255737	1993	9373	7698.370387	2001	11666	11408.448014
1986	2200	2238.022389	1994	10143	8472.663076	2002	10122	11558.462675
1987	1850	2821.459595	1995	10327	9159.866955	2003	10355	11672.955439
1988	2861	3502.548747	1996	11328	9749.725874	2004	9168	11759.845412

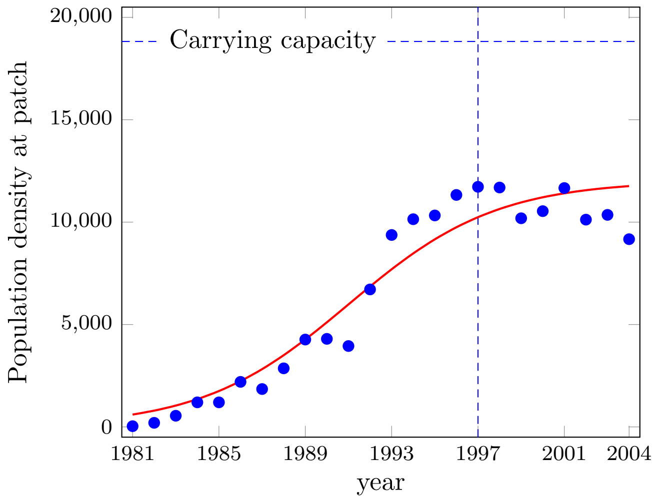


Figure S17. Upper: Predictions for the period 1981–2004 obtained from the best fit of Equation (3.1) to the field data taking the value β estimated from the initial phase and leaving free $x(0)$ and α.
Lower: Dynamics obtained for the best fitting (red line). Field data shown with blue dots. Notice that the predicted initial condition largely departs from the observed one. Here we obtained LS= 5639.340356 and $R^2 = 0.9277$.

$$\frac{\mathrm{d}}{\mathrm{d}t} y(t) = \varphi y(t) - \beta y(t)^2 - \lambda D(y(t), \mu, \sigma, \delta) \tag{4.5}$$

with the following parameters

Parameter	**Range or value**	**Ecological meaning or description**
α	0.3489494104672237	Population growth rate including death of individuals (without linear dispersal)
β	$2.43826356697 \times 10^{-5}$	Intrinsic growth rate over the carrying capacity
K	18822.79734	Carrying capacity
$y(0)$	$x(16) = 12096.688846$	Initial condition set to the 1997 population size estimated from the initial phase fitting
μ	0.0	Tendency of dispersal function for small population sizes
σ	0.4904756364357690	Sharpness and smoothness of the dispersal function
δ	8944.228275	Transition between small and large population sizes
ρ	$\mathbb{R}^+$	Linear (exponential) dispersal rate
$\varphi = \alpha - \rho$	$(-\infty, \alpha]$	Population growth rate including linear dispersal
λ	$\mathbb{R}^+$	Dispersal rate

The solution of the above Model (4.5) with initial condition $x(16) = x(1997) = 12096.688846$ will be denoted by $y(t) = y_{\varphi,\lambda}(t)$. Obviously, $y_{\varphi,\lambda}(0) = x(16) = 12096.688848$.

Observe that the fitting in this setting will consist in estimating two parameters: the population growth rate including a possible linear dispersal φ, and the dispersal rate λ. In a similar way as before we define the map

$$\begin{aligned} \mathsf{T}\colon \quad [-\infty,\alpha]\times\mathbb{R}^{+} &\longrightarrow \mathbb{R}^{+} \\ (\varphi,\lambda) &\longmapsto \sqrt{\textstyle\sum_{\ell=1}^{7}\big(y_{\varphi,\lambda}(\ell)-\psi(\ell)\big)^{2}}, \end{aligned}$$

where

$$\psi(\ell,\ell=1:7) = \mathsf{Audouin's\_Gulls\_Observed\_Population\_at\_year}(1997+\ell,\ell=1:7) = \big[\overset{[1]}{11691},\overset{[2]}{10189},\overset{[3]}{10537},\overset{[4]}{11666},\overset{[5]}{10122},\overset{[6]}{10355},\overset{[7]}{9168}\big].$$

Now, the *fitting of the model* consists in solving

$$\begin{aligned} &\min \mathsf{T}\big(\varphi,\lambda\big) \\ &\text{subject to } \big(\varphi,\lambda\big)\in[-\infty,\alpha]\times\mathbb{R}^{+}, \\ &\text{and } y(t)\in[0,K] \text{ for } t\in[0,7]. \end{aligned} \tag{4.6}$$

To solve this problem again we have used a standard *trust region method* with the Levenberg-Marquardt algorithm to solve the trust region sub-problem (see the GNU Scientific Library (GSL) *Nonlinear Least-Squares Fitting* documentation). As before, we have used numerical approximation of derivatives of the objective function. The obtained results are given by $\varphi = 0.2983089056790262$ and $\lambda = 1352.9233$, with error $\mathsf{T}\big(\varphi,\lambda\big) = 1675.012608\cdots$. Despite the low amount of data, these parameters also indicate that the dominant dispersal is due to social copying, since we obtained $\rho = 0.050640505$, meaning that the positive density-dependent dispersal contribution is extremely low as compared to the social copying behaviour for dispersal. The model fitting for this short period of time is shown in Figure S18 below.

Year	Population Data	
	Observed	Predicted
1997	11725	12096.688846
1998	11691	11646.563095
1999	10189	11258.159411
2000	10537	10908.587887
2001	11666	10579.827439
2002	10122	10255.190485
2003	10355	9915.362227
2004	9168	9530.934270

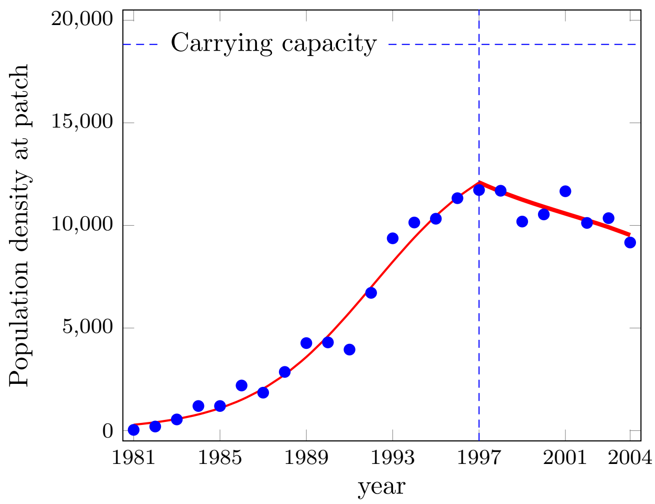


Figure S18. Left: Model fitting for the period 1998–2004 obtained using Equation (3.1) taking the value β estimated from the initial phase. Here, we have left free φ and λ, using the values of μ, σ, and δ obtained with the best fit for the collapse phase ($\mu = 0.0$, $\sigma = 0.4904756364357690$ and $\delta = 8944.228275$). We have assumed that the shape of the function $D(x,\mu,\sigma,\delta)$ is a trait of this species.
Right: Dynamics obtained for the best fitting. The thin red line shows the fitting of the initial phase performed in Section 3.2.2. The fitting of the dynamics from 1998 to 2004 (thick red line) results in a quadratic error $\mathsf{LS} = 1675.012608$ and $\mathsf{LS} = 3086.943892729$ for the period 1981–2004. Field data is shown with blue dots.

Supplementary Section 5

Alternative dispersal models to social copying fail to fit the collapse phase 2006–2017

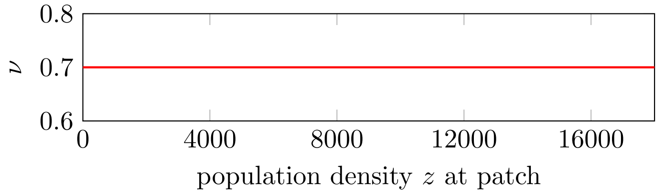
In this section, we will use two alternative dispersal modes considering explicitly density-independent dispersal and positive density-dependent dispersal to test whether these modes are able to improve the fit of the collapse phase obtained with Equation (4.1). We note that the fitting of the full collapse phase performed in Section 4.2 included a mode of positive density-dependent dispersal (by means of parameter ρ included in φ) and the Elliot sigmoid function, which, for some parameter combinations can behave as a density-independent function (see e.g., red line in the panel at the right in Figure S7, Figure S10 and Remark 4.2). The results obtained for the best fit indicated that the contribution of positive density-dependent dispersal was extremely low compared to the social copying mode ($\rho \approx 0.03$). Also, the sigmoid function obtained was far from a density-independent dispersal mode (see the graphs in Page 38, and Remark 4.19). However, in order to provide further evidence of the social copying mechanism explaining the data for the full collapse phase we will here investigate two models with these alternative dispersal modes.

5.1 Model fit with explicit density-independent dispersal

In this subsection we use the following model with *density-independent dispersal* to fit the period from 2006 to 2017:

$$\frac{\mathrm{d}}{\mathrm{d}t}\, z(t) = \gamma z(t)\left(1 - \frac{z(t)}{K}\right) - \varepsilon z(t) - \nu = \alpha z(t) - \beta z(t)^2 - \nu. \tag{5.1}$$

Here, ν denotes a constant tendency to dispersal, which is independent on the population density (see the plot below for $\nu = 0.7$), and the parameters $\alpha = \gamma - \varepsilon$ and $\beta = \frac{\gamma}{K}$ are taken equal to the ones computed in Section 3.2.2 (recalled in the table below), since these values estimated from the initial phase are considered as intrinsic population's characteristics. Notice that here we do not include the term $-\rho z(t)$ to consider only density-independence in dispersal.



The solution of Model (5.1) will be denoted by $z(t) = z_\nu(t)$, $t \in [0, 11]$. As usual, it depends on the initial condition $z(0) = z_\nu(0) \in [0, K]$, that must be considered a free parameter as well.

The parameters' fitting of Model (5.1) will be done with the help of the map

$$\begin{array}{rccc} \mathsf{DI}\colon & [0,K] \times \mathbb{R}^+ & \longrightarrow & \mathbb{R}^+ \\ & \big(z(0), \nu\big) & \longmapsto & \sqrt{\sum_{\ell=0}^{11} \big(z_\nu(\ell) - \pi(\ell)\big)^2}, \end{array}$$

and by considering the *denso-independent minimization problem*

$$\begin{aligned} &\min \ \mathsf{DI}\big(z(0),\nu\big) \\ &\text{subject to } \nu \in \mathbb{R}^+, \\ &\text{and } z(t) \in [0,K] \text{ for } t \in [0,11]. \end{aligned} \tag{5.2}$$

As before, Equation (5.1) is a Ricatti equation with constant coefficients, but this time with (non-zero) independent term. The next lemma gives analytic expressions for $z(t)$, and hence, for $\mathsf{DI}\big(z(0),\nu\big)$.

Lemma 5.1 (A Ricatti Equation with constant coefficients and independent term)**.** *Assume that in Equation (5.1) α, β and ν are positive. Let $\xi = 4\beta\nu - \alpha^2$. Then, the solution $z(t)$ of Equation (5.1) is the following.*

(a) If $\xi > 0$, then

$$z(t) = \frac{(z(0)\alpha - 2\nu)\tan\big(\frac{\sqrt{\xi}}{2}t\big) + z(0)\sqrt{\xi}}{(2z(0)\beta - \alpha)\tan\big(\frac{\sqrt{\xi}}{2}t\big) + \sqrt{\xi}}.$$

(b) If $\xi < 0$, then

$$z(t) = \frac{\big(z(0)\left(\sqrt{-\xi}+\alpha\right) - 2\nu\big)\exp\big(\sqrt{-\xi}t\big) + z(0)\left(\sqrt{-\xi}-\alpha\right) + 2\nu}{\big(\sqrt{-\xi} + 2z(0)\beta - \alpha\big)\exp\big(\sqrt{-\xi}t\big) + \sqrt{-\xi} + \alpha - 2z(0)\beta}.$$

(c) When $\xi = 0$,

$$z(t) = \frac{\alpha(2z(0)\beta - \alpha)t + 4z(0)\beta}{2\beta(2z(0)\beta - \alpha)t + 4\beta}.$$

As an exploratory step, Figure S19 below shows the graph of $\mathsf{DI}\big(\pi(0),\nu\big)$ with $\nu \in [0, 10000]$. The most interesting thing from the data plot is the fact that $\mathsf{DI}\big(\pi(0), 2243.8\big) = 3725.6611\cdots$.

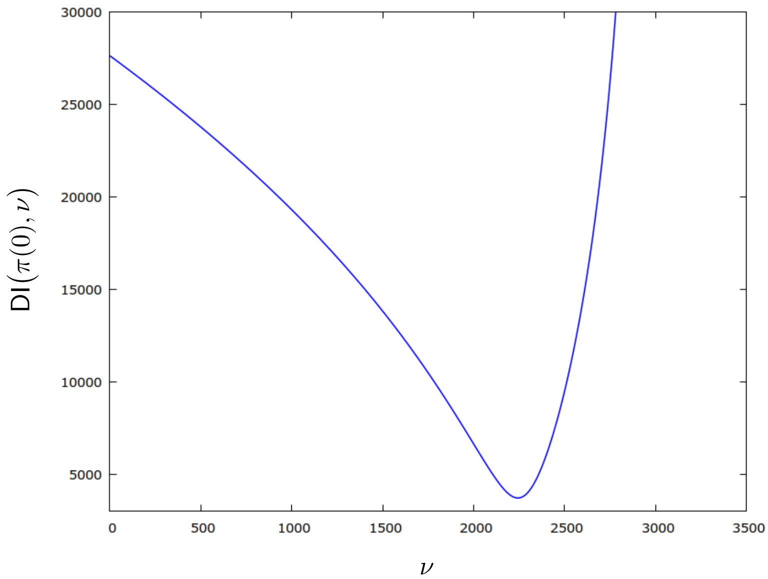


Figure S19. Graph of $\mathsf{DI}\big(\pi(0),\nu\big)$ with $\nu \in [0, 10000]$. Some distinguished values of the function $\mathsf{DI}\big(\pi(0),\nu\big)$ are:

- $\mathsf{DI}\big(\pi(0), 0\big) = 27621.23\cdots$,
- $\mathsf{DI}\big(\pi(0), 2781.8\big) = 30001.3433\ldots$,
- $\mathsf{DI}\big(\pi(0), 3500\big) = 163438.632234\ldots$, and
- $\mathsf{DI}\big(\pi(0), 2243.8\big) = 3725.6611\cdots$.

Now we will use the fact that $\mathsf{DI}\big(\pi(0), 2243.8\big) = 3725.6611\cdots$ to narrow the parameters' search space for Problem (5.2).

Lemma 5.2. *We have:*

$$\underset{\vec{\boldsymbol{\theta}} \in [0,K]\times\mathbb{R}^+}{\arg\min} \ \mathsf{DI}\big(\vec{\boldsymbol{\theta}}\big) \in (11603, K] \times (1800, 2410).$$

Remark 5.3. When $\nu > 1800$, by using the notation of Lemma 5.1, we have

$$\xi = 4\beta\nu - \alpha^2 > 7200\beta - \alpha^2 \approx 0.053789 > 0.$$

Thus, the solution $z(t) = z_\nu(t)$ of Model (5.1) for $\big(z_\nu(0), \nu\big) \in (11603, K] \times (1800, 2410)$ is the one from Lemma 5.1(a).

Proof of Lemma 5.2. We will use again the ideas from the proof of Lemma 4.11. Assume that

$$\big(\kappa^*, \nu^*\big) = \underset{\vec{\boldsymbol{\theta}} \in [0,K]\times\mathbb{R}^+}{\arg\min} \ \mathsf{DI}\big(\vec{\boldsymbol{\theta}}\big),$$

and $z_{\nu^*}(0) = \kappa^* < 11603 = \pi(0) - 3726$. Then,

$$\mathsf{DI}\big(\kappa^*, \nu^*\big) = \sqrt{\sum_{\ell=0}^{11} \big(z_{\nu^*}(\ell) - \pi(\ell)\big)^2} \geq |z_{\nu^*}(0) - \pi(0)| > 3726 > \mathsf{DI}\big(\pi(0), 2243.8\big);$$

a contradiction.

Next we denote by $v(t)$, $t \in [0, 11]$, the solution of

$$\frac{dv(t)}{dt} = \alpha v(t) - \beta v(t)^2 - 1800,$$

with initial condition $v(0) = 11603$. By direct computation, we get

$$\begin{array}{llll} v(8) & \approx 6310.75 & > & \pi(8) \ = 4778, \\ v(9) & \approx 5727.19 & > & \pi(9) \ = 2067, \\ v(10) & \approx 5101.31 & > & \pi(10) = 1586, \\ v(11) & \approx 4409.11 & > & \pi(11) = \ \ 793, \end{array}$$

(in particular $v(t)$ is defined, non-negative and bounded on the interval $[0, 11]$), and

$$\sqrt{\sum_{\ell=8}^{11} \big(v(\ell) - \pi(\ell)\big)^2} = 6417.1639\cdots.$$

Observe that when $\nu^* \leq 1800$,

$$\alpha x - \beta x^2 - 1800 \leq \alpha x - \beta x^2 - \nu^*$$

for every $x \in \mathbb{R}^+$. Moreover, since $z_{\nu^*}(0) > 11603 = v(0)$, by Lemma 4.13(a),

$$z_{\nu^*}(\ell) \geq v(\ell) > \pi(\ell)$$

for $\ell = 8, 9, 10, 11$. Hence,

$$\mathsf{DI}\big(z_{\nu^*}(0), \nu^*\big) \geq \sqrt{\sum_{\ell=8}^{11} \big(z_{\nu^*}(\ell) - \pi(\ell)\big)^2} \geq \sqrt{\sum_{\ell=8}^{11} \big(v(\ell) - \pi(\ell)\big)^2} > 6417 > \mathsf{DI}\big(\pi(0), 2243.8\big);$$

again a contradiction.

Now we will show that whenever $\nu^* \geq 2410$, the solution $z_{\nu^*}(t)$ with $z_{\nu^*}(0) \in (11603, K]$ is not feasible (i.e. $z_{\nu^*}(\ell) < 0$ for some $\ell \in \{1, 2, \ldots, 11\}$). To do this we assume by way of contradiction that $z_{\nu^*}(t)$ is defined (and non-negative) in the interval $[0, 11]$. On the other hand, the solution of

$$\frac{dx(t)}{dt} = \alpha x(t) - \beta x(t)^2 - 2410,$$

with initial condition $\widehat{v}(0) = 18823 > K \geq z_{\nu^*}(0)$ will be denoted by $\widehat{v}(t)$. By direct computation we obtain $\widehat{v}(11) < -160$ (and $\widehat{v}(t) > 1949$ for $t \in [0, 10]$).

Since $\nu^* \geq 2410$ and $z_{\nu^*}(0) < \widehat{v}(0)$, we get that

$$\alpha x - \beta x^2 - \nu^* \leq \alpha x - \beta x^2 - 2410,$$

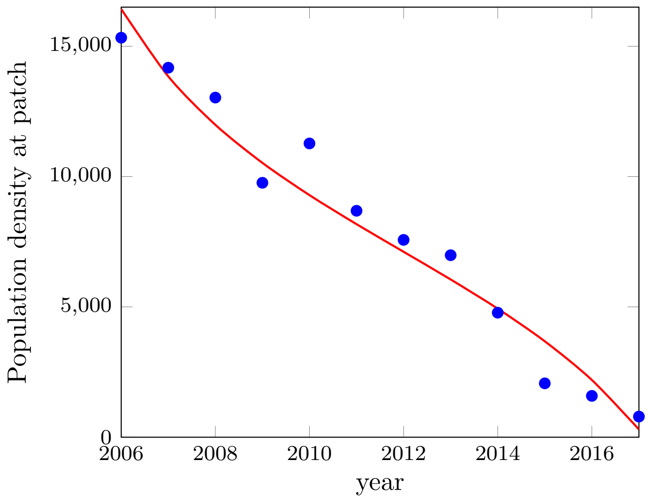
and $0 \leq z_{\nu^*}(11) \leq \widehat{v}(11) < -160$; a contradiction. This ends the proof of the lemma. $\square$

By using again a trust region method with the Levenberg-Marquardt algorithm with numerical approximation of derivatives of the objective function we have obtained the solution of Problem (5.2) described in the table below:

Parameters' values at the optimum	
Parameter	**Value**
α	0.3489494104672237
β	$2.43826356697 \times 10^{-5}$
$z(0)$	$16428.1638 \cdots$
ν	$2300.1381 \cdots$
Quadratic Error $= \mathsf{DI}(z(0), \nu)$	$3391.962 \cdots$
Coefficient of determination R^2	$0.95759 \cdots$

and the prediction generated by the solution of Problem (5.2) is listed and displayed below:

Year	**Population Data**	
	Observed	**Predicted**
2006	15329	16428.16
2007	14177	13842.56
2008	13031	11987.27
2009	9762	10526.06
2010	11271	9288.54
2011	8688	8174.37
2012	7571	7114.67
2013	6983	6052.82
2014	4778	4931.77
2015	2067	3680.99
2016	1586	2196.48
2017	793	299.55



However, since the trust method with the Levenberg-Marquardt algorithm is extremely sensitive to the choice of the initial seed we also have implemented a non efficient but extremely robust method based on the Brent one-dimensional search algorithm (see again the GNU Scientific Library (GSL) *One Dimensional Minimization* documentation). The Brent one-dimensional search algorithm is extremely robust since it requires as starting information a *minimum bracketing interval* and gives the solution inside a small (depending on the tolerance) minimum bracketing interval, thus guaranteeing that the minimum in search exists and will not be missed during the search. A *minimum bracketing interval* for a function G is a triplet of points in the real line $a < m < b$ such that $G(a) > G(m) < G(b)$. Of course, in these conditions, it is guaranteed that the function G has a minimum in the interval (a, b).

How is it possible to use a one-dimensional search algorithm in computing a minimum in two variables (initial condition and ν)? We use a method that we call a *re-iterative one-dimensional search by the Brent minimization algorithm*, and we perform it with the help of the following one-parameter auxiliary family of maps with parameter $\kappa \in [11603, K]$:

$$\begin{aligned} \Delta_\kappa\colon \quad [1800, 2410] &\longrightarrow \mathbb{R}^+ \\ \nu &\longmapsto \mathsf{DI}(\kappa, \nu) \end{aligned}$$

Observe that, for every $\kappa \in [11603, K]$, Δ_κ has a minimum in the interval $(1800, 2410)$ by Lemma 5.2. In this setting it is also helpful to consider the one-dimensional function

$$\begin{aligned} \Psi\colon \quad [11603, K] &\longrightarrow \mathbb{R}^+ \\ \kappa &\longmapsto \mathsf{DI}\left(\kappa, \operatorname*{arg\,min}_{\nu \in [1800,2410]} \Delta_\kappa(\nu)\right) \end{aligned}$$

As for the maps Δ_κ, Lemma 5.2 and direct computations tell us that $11603 < 15329 < K$ is always a minimum bracketing interval for the map Ψ.

Then, again by Lemma 5.2,

$$\min_{\substack{\kappa\in[0,K]\\ \nu\in\mathbb{R}^+}} \mathsf{DI}\big(\kappa,\nu\big) = \min_{\substack{\kappa\in[11603,K]\\ \nu\in[1800,2410]}} \mathsf{DI}\big(\kappa,\nu\big) = \min_{\kappa\in[11603,K]} \mathsf{DI}\left(\kappa, \operatorname*{arg\,min}_{\nu\in[1800,2410]} \Delta_\kappa(\nu)\right) = \min_{\kappa\in[11603,K]} \Psi(\kappa)$$

which, by using Brent minimization algorithm, gives an extremely robust method to re-iteratively compute the solution of Problem (5.2). As we have already said, the minimum bracketing interval for the map Ψ (i.e. for the first minimization) is known to be $11603 < 15329 < K$. Unfortunately, the minimum bracketing interval for the maps Δ_κ are not uniform and depend on κ. More precisely, they must be of the form $1800 < m_\kappa < 2410$ (of course with m_κ depending on κ). The value of m_κ is determined for each value of κ with an iterative procedure similar to bisection but choosing m_κ from meshes of the form $1800 + \frac{i}{2^k} \cdot (24100 - 1800)$ with $i \in \{1, 2, \dots, 2^k - 1\}$, and k small ($k < 9$ is sufficient).

Fortunately, the *re-iterative one-dimensional search by the Brent minimization algorithm* gives the same solution that the trust method with the Levenberg-Marquardt algorithm described in the above two tables and figure.

5.2 Model fit with positive density-dependent dispersal

The model used here also considers the structural parameters, together with a dispersal function obtained by reversing the D function defined to model social copying, thus having positive density-dependent dispersal (i.e., the larger the population size at patch the stronger the dispersal). This model is given by:

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}t}x(t) &= \gamma x(t)\left(1 - \frac{x(t)}{K}\right) - \varepsilon x(t) - \rho x(t) - \lambda P\big(x(t),\mu,\sigma,\delta\big)\\ &= \varphi x(t) - \beta x(t)^2 \quad - \underbrace{\boxed{\lambda P\big(x(t),\mu,\sigma,\delta\big),}}_{\substack{\text{Positive density-dependent}\\ \text{dispersal term}}}\end{aligned} \tag{5.3}$$

now with

$$\boxed{P(x,\mu,\sigma,\delta) := 1 - D(x,\mu,\sigma,\delta),} \tag{5.4}$$

and the parameters as described in the table below. Observe that here we have chosen $\mu \geq 1$ in order that $P(x,\mu,\sigma,\delta)$ is increasing with respect to x.

Parameter	Range or value	Ecological meaning or description
$\alpha = \gamma - \varepsilon$	0.3489494104672237	Population growth rate including death of individuals (without linear dispersal)
ρ	$\mathbb{R}^+$	Linear (exponential) dispersal rate
$p(0)$	$[0, K]$	Initial condition
$\varphi = \alpha - \rho$	$(-\infty, \alpha]$	Neat population growth rate including linear dispersal
$\beta = \frac{\gamma}{K}$	$2.43826356697 \times 10^{-5}$	Intrinsic growth rate over the carrying capacity
Parameters concerning the model positive density-dependent dispersal rate		
λ	$\mathbb{R}^+$	Dispersal rate
μ	$[1, +\infty)$	Tendency of dispersal function for small population sizes
σ	$\mathbb{R}^+$	Sharpness and smoothness of the dispersal function
δ	$\mathbb{R}^+$	Transition between small and large population sizes

Note that, similarly to Model (4.1), we are here considering a continuum of functions now obeying a positive-density dependent dispersal rate (the more the individuals at the patch, the stronger the dispersal), given by Equation (5.4). Figure S20 below displays the shape of the Function (5.4) above for several range of parameters, and shows the possible different nonlinear behaviours in terms of parameters.

The solution of Model (5.3) with parameters belonging to the ranges shown in the above table will be denoted by

$$p(t) = p_{\varphi,\lambda,\mu,\sigma,\delta}(t), \qquad t \in [0, 11].$$

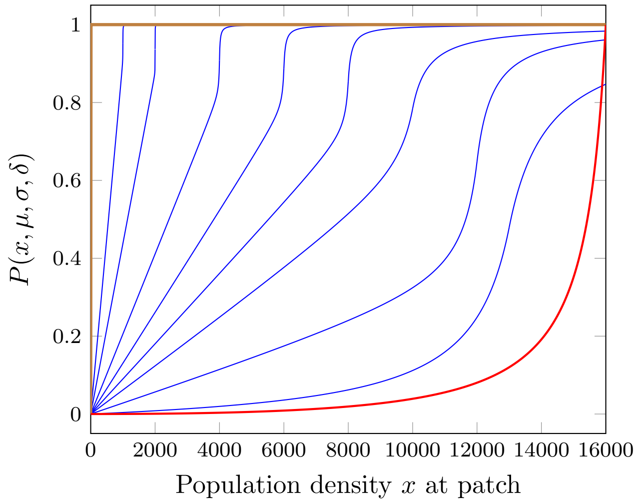


Figure S20. Several examples of the function $P(x, \mu, \sigma, \delta)$ for a population size $x \in [0, 16000]$, and parameter ranges $\mu \in [1, 100]$, $\sigma \in [0.75, 1000]$, $\delta \in [10, 15335]$. For instance, the lower curve has been obtained with $\mu = 1$, $\sigma = 0.75$, and $\delta = 15335$ (red line); while the Heaviside-like brown curve has been plotted with $\mu = 100$, $\sigma = 500$, and $\delta = 10$.

The solution $p(t)$ depends on the initial condition $p(0) = p_{\varphi,\lambda,\mu,\sigma,\delta}(0) \in [0, K]$, that is another free parameter. On the other hand, recall that we have denoted the observed population of Audouin's gulls at the years 2006 to 2017 by

$$\pi(\ell, \ell = 0:11) = \textsf{Audouin's\_Gulls\_Observed\_Population\_at\_year}(2006 + \ell, \ell = 0:11) =$$
$$\big[\overset{[0]}{15329}, \overset{[1]}{14177}, \overset{[2]}{13031}, \overset{[3]}{9762}, \overset{[4]}{11271}, \overset{[5]}{8688}, \overset{[6]}{7571}, \overset{[7]}{6983}, \overset{[8]}{4778}, \overset{[9]}{2067}, \overset{[10]}{1586}, \overset{[11]}{793}\big].$$

and we have defined the parameter space $\mathscr{P} := [0, K] \times [-\infty, \alpha] \times \mathbb{R}^+ \times \mathbb{R}^+ \times \mathbb{R}^+ \times \mathbb{R}^+$.

Now we define the map

$$\begin{array}{rccc} \textsf{PDD}\colon & \mathscr{P} & \longrightarrow & \mathbb{R}^+ \\ & \big(p(0), \varphi, \lambda, \mu, \sigma, \delta\big) & \longmapsto & \sqrt{\sum_{t=0}^{11} \big(p(t) - \pi(t)\big)^2}. \end{array}$$

Again, the *fitting of the model with positive density-dependent dispersal* consists in solving

$$\begin{aligned} &\min \textsf{PDD}\big(\kappa, \varphi, \lambda, \mu, \sigma, \delta\big) \\ &\text{subject to } \big(\kappa, \varphi, \lambda, \mu, \sigma, \delta\big) \in \mathscr{P}, \\ &\quad p_{\varphi,\lambda,\mu,\sigma,\delta}(0) = \kappa, \\ &\text{and } p_{\varphi,\lambda,\mu,\sigma,\delta}(t) \in [0, K] \text{ for } t \in [0, 11], \end{aligned} \tag{5.5}$$

and checking that this minimum is as low as possible to guarantee the validity of the model.

As before we start with a brute force exploration of a reasonable region of the parameter space $\mathscr{P}$ to get a reasonable upper bound of the optimum Quadratic Error value that will be used as a fulcrum to analytically determine a good search region for the genetic algorithm. The brute force exploration in this case has been implemented as a Sparse Anisotropic Grid Search on a compact sub-region of $\mathscr{P}$, and a second (finer and isotropic) grid search in a small sub-region of the Sparse Anisotropic Grid Search region to improve the previously obtained brute-force best fit. The parameters and meshes for both grid searches are completely specified in Table 5.The results obtained are summarized in the next lemma.

Lemma 5.4. *We have*

$$\begin{aligned} \textsf{PDD}\big(16400, \alpha, 2300, 10, 400, 1400\big) &= 3354.0912875227850\cdots \quad \textit{and} \\ \textsf{PDD}\big(16522, \alpha, 2330, 8, 450, 1495\big) &= 3322.5872893066994038\cdots. \end{aligned}$$

$p_{(\alpha,2330,8,450,1495)}(0) = 16519$, *and* $p_{(\alpha,2330,8,450,1495)}(t) \in [0, K]$ *for* $t \in [0, 11]$.

Next we will use the above lemma to analytically determine a better (concerning the search of the solution of Problem 5.5) subdomain of $\mathscr{P}$.

Parameter	Theoretical Range	Sparse Anisotropic Grid Search: Effective Search Range	Sparse Anisotropic Grid Search: Anisotropic Step	Fine Grid Search: Effective Search Range	Fine Grid Search: Step
$p_{\varphi,\nu,\lambda,\mu,\sigma,\delta}(0)$	$[0, K]$	$[10000, 19000]$	200	$[15329, 16600]$	1
φ	$(-\infty, \alpha]$	$[0, \alpha]$	0.01	$[0.3, \alpha]$	0.01
λ	$\mathbb{R}^+$	$[0, 3000]$	100	$[2200, 2400]$	10
μ	$\mathbb{R}^+$	$[1, 10]$	$\begin{cases} 0.1 & \text{when } \mu \in [1, 1.9], \\ 1 & \text{when } \mu \in [2, 10] \end{cases}$	$[8, 12]$	0.1
σ	$\mathbb{R}^+$	$[0, 400]$	$\begin{cases} 1 & \text{when } \sigma \in [0, 9], \\ 5 & \text{when } \sigma \in [10, 45], \\ 10 & \text{when } \sigma \in [50, 400] \end{cases}$	$[330, 450]$	10
δ	$\mathbb{R}^+$	$[0, 20000]$	100	$[100, 1500]$	5

Table 5. Full specification of the Sparse Anisotropic Grid Search for the model with positive density-dependent dispersal. The SAGS has explored (computed the fitness for) 5,120,766,000 mesh points while the Fine Grid Search has explored 19,932,494,400 mesh points.

Lemma 5.5. *Let* $(\kappa^*, \varphi^*, \lambda^*, \mu^*, \sigma^*, \delta^*) \in \mathscr{P}$ *be a minimum of Problem* (5.5). *Then,*

$$\mathsf{PDD}(\kappa^*, \varphi^*, \lambda^*, \mu^*, \sigma^*, \delta^*) \leq \mathsf{PDD}(16522, \alpha, 2330, 8, 450, 1495) = 3322.5872893066994038\cdots,$$

and $\kappa^* \in [12006, 18652]$, $\varphi^* \in [0.1, \alpha]$, *and* $\lambda^* > 400$.

Proof. The first statement of the lemma is obvious. The proof that $p_{\varphi^*,\lambda^*,\mu^*,\sigma^*,\delta^*}(0) = \kappa^* \in [12006, 18652]$ is analogous to the proof of the bound for the initial condition in Lemma 5.2.

Now we assume by way of contradiction that $\varphi^* < 0.1$. Then, as in Lemma 4.11, we denote by $u(t)$, $t \in [0, 11]$ the solution of Model (3.1) with α replaced by $\widetilde{\varphi} = 0.1$, $\beta = 2.43826356697 \times 10^{-5}$, and initial condition $u(0) = 18652$. Since

$$u(t) = \frac{\widetilde{\varphi} u(0) \exp(\widetilde{\varphi} t)}{\widetilde{\varphi} + \beta u(0)(\exp(\widetilde{\varphi} t) - 1)},$$

by direct computation we get that $u(t)$ is defined, non-negative and bounded on the interval $[0, 11]$,

$$\begin{array}{lcl}
u(1) \approx 13944.14\cdots & < & \pi(1) = 14177, \\
u(2) \approx 11351.60\cdots & < & \pi(2) = 13031, \\
u(3) \approx \ \ 9716.92\cdots & < & \pi(3) = \ \ 9762, \\
u(4) \approx \ \ 8596.75\cdots & < & \pi(4) = 11271, \\
u(5) \approx \ \ 7784.73\cdots & < & \pi(5) = \ \ 8688, \\
u(6) \approx \ \ 7171.78\cdots & < & \pi(6) = \ \ 7571, \\
u(7) \approx \ \ 6694.80\cdots & < & \pi(7) = \ \ 6983,
\end{array}$$

and

$$\sqrt{\sum_{\ell=1}^{7} \big(\pi(\ell) - u(\ell)\big)^2} = 3329.639782820\cdots.$$

Since $\mu^* \geq 1$, by Proposition 4.1(a-c), $D(x, \mu^*, \sigma^*, \delta^*) \in (0, 1]$. Hence $1 - D(x, \mu^*, \sigma^*, \delta^*) \geq 0$, and consequently,

$$\varphi^* x - \beta x^2 - \lambda^*\big(1 - D(x, \mu^*, \sigma^*, \delta^*)\big) \leq \widetilde{\varphi} x - \beta x^2,$$

for every $x \in \mathbb{R}^+$. So, since $p(0) \leq 18652 = u(0)$, we get by Lemma 4.13(b) that either $p(t)$ is not defined for every t in the interval $[0, 11]$ (in particular $p(t)$ is not feasible), or

$$p(\ell) \leq u(\ell) < \pi(\ell)$$

for $\ell = 1, 2, \ldots, 7$. Hence,

$$\mathsf{PDD}\big(\kappa^*,\varphi^*,\lambda^*,\mu^*,\sigma^*,\delta^*\big) \geq \sqrt{\sum_{\ell=1}^{7}\big(\pi(\ell)-p(\ell)\big)^2} \geq$$
$$\sqrt{\sum_{\ell=1}^{7}\big(\pi(\ell)-u(\ell)\big)^2} > 3329 > \mathsf{PDD}\big(16522,\alpha,2330,8,450,1495\big);$$

a contradiction.

If $\lambda^* \leq 400$, $\lambda^*\big(1-D\big(x,\mu^*,\sigma^*,\delta^*\big)\big) < \lambda \leq 400$ because $D\big(x,\mu^*,\sigma^*,\delta^*\big) \in (0,1]$. Thus,

$$0.1x-\beta x^2-400 \leq \varphi^* x-\beta x^2-\lambda^* P\big(x,\mu^*,\sigma^*,\delta^*\big)$$

for every $x \in \mathbb{R}^+$. Next we denote by $z(t)$, $t \in [0,11]$ the solution of

$$\frac{dx(t)}{dt} = 0.1z(t)-\beta z(t)^2-400, \tag{5.6}$$

with initial condition $z(0) = 12006$. by direct computation we get that $z(t)$ is defined, non-negative and bounded on the interval $[0,11]$,

$$\begin{array}{lcl} z(9) \approx 3811.19\cdots & > & \pi(9) = 2067, \\ z(10) \approx 3452.67\cdots & > & \pi(10) = 1586, \\ z(11) \approx 3117.91\cdots & > & \pi(11) = \ \ 793, \end{array}$$

and

$$\sqrt{\sum_{\ell\in\{9,10,11\}}\big(\pi(\ell)-z(\ell)\big)^2} = 3454.2648412690\cdots .$$

Then, since $p(0) > 12006 = z(0)$, by Lemma 4.13(a),

$$p(\ell) \geq z(\ell) > \pi(\ell)$$

for $\ell = 9,10,11$. Hence,

$$\mathsf{PDD}\big(\kappa^*,\varphi^*,\lambda^*,\mu^*,\sigma^*,\delta^*\big) \geq \sqrt{\sum_{\ell\in\{9,10,11\}}\big(p(\ell)-\pi(\ell)\big)^2} \geq$$
$$\sqrt{\sum_{\ell\in\{9,10,11\}}\big(z(\ell)-\pi(\ell)\big)^2} > 3454 > \mathsf{PDD}\big(16522,\alpha,2330,8,450,1495\big);$$

again a contradiction. □

Numerical simulations complementing the two grid searches performed for Model (5.3) with Function (5.4) give the following results:

Numerical Result 1. *The following statements hold:*

(a) For every $\mu \geq 700$ and $\sigma \geq 6000$ fixed, there exists $\delta^* \approx 338.8716$ such that

$$\min_{\substack{\kappa\in[12006,18652]\\ \varphi\in[0.1,\alpha]\\ \lambda\in(400,\infty)\\ \delta\in\mathbb{R}^+}} \mathsf{PDD}\big(\kappa,\varphi,\lambda,\mu,\sigma,\delta\big) = \mathsf{PDD}\big(16428.163730,\alpha,2300.1381396,\mu,\sigma,\delta^*\big).$$

Moreover, $p_{\alpha,2328.9,\mu,\sigma,\delta^*}(t) \in [0,K]$ for $t \in [0,11]$. That is the optimal solutions of Problem (5.5) for $\mu \geq 700$ and $\sigma \geq 6000$ are feasible.

(b) There exists a value $\Xi = 3391.484550\cdots$ such that

$$\lim_{\mu,\sigma\to\infty}\min_{\substack{\kappa\in[12006,18652]\\ \varphi\in[0.1,\alpha]\\ \lambda\in(400,\infty)\\ \delta\in\mathbb{R}^+}} \mathsf{PDD}\big(\kappa,\varphi,\lambda,\mu,\sigma,\delta\big) = \lim_{\mu,\sigma\to\infty}\mathsf{PDD}\big(16428.163730,\alpha,2300.1381396,\mu,\sigma,\delta^*\big) = \Xi.$$

Numerical Result 2.

$$\min_{\substack{\kappa\in[12006,18652]\\ \varphi\in[0.1,\alpha]\\ \lambda\in(400,\infty)\\ \mu\in[1,\infty)\\ \sigma\in\mathbb{R}^+\\ \delta\in\mathbb{R}^+}} \mathsf{PDD}\big(\kappa,\varphi,\lambda,\mu,\sigma,\delta\big) = \min_{\substack{\mu\in[1,\infty)\\ \sigma\in\mathbb{R}^+}} \min_{\substack{\kappa\in[12006,18652]\\ \varphi\in[0.1,\alpha]\\ \lambda\in(400,\infty)\\ \delta\in\mathbb{R}^+}} \mathsf{PDD}\big(\kappa,\varphi,\lambda,\mu,\sigma,\delta\big) > 3300.$$

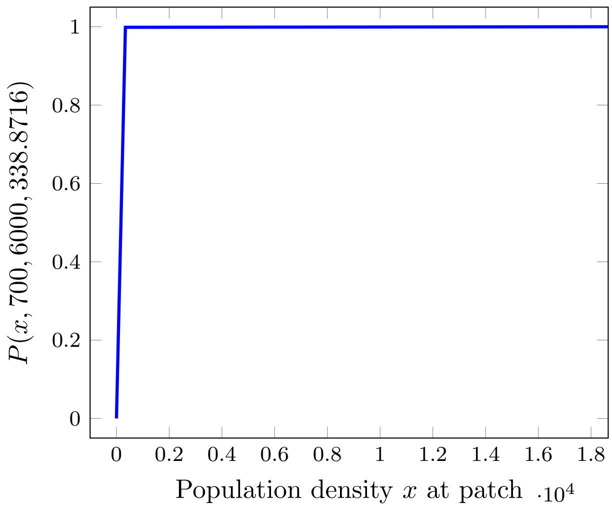


Figure S21. The graph of the function $P(x,\mu,\sigma,\delta)$ for a population size $x \in [0, 18652]$ and $\mu = 700$, $\sigma = 6000$ and $\delta = 338.8716$. Note that $P(338.8716, 700, 6000, 338.8716) = 0.998572065260\ldots$, and $P(18652, 700, 6000, 338.8716) = 0.99999998700\ldots$. The graphs of $P(x,\mu,\sigma,338.8716)$ with $\mu \geq 700$ and $\sigma \geq 6000$ are qualitatively indistinguishable from the one shown here.

As a conclusion we remark that if a minimum of Problem (5.5) exists, it must verify either $\mu < 700$ or $\sigma < 6000$, and its associated fitness value must be larger than 3300. Therefore, *the fitness (least squares approximation to the data) of any model with positive density-dependent dispersal is far from the one obtained with the model with a negative density-dependent dispersal.* Consequently, the positive density-dependent dispersals must be discarded as explanatory models in favour of the ones with negative density-dependent dispersal

Bibliography

[1] Bartumeus, F. *et al.* Fishery Discards Impact on Seabird Movement Patterns at Regional Scales. Curr. Biol. 20, 215–222 (2010).

[2] Oro, D., Ruiz, X. Seabirds and trawler fisheries in the northwestern Mediterranean: differences between the Ebro Delta and the Balearic Is. areas. ICES J. Mar. Sci. 54, 695–707 (1997).

[3] Martínez-Abraín, A., Oro, D., Forero, M. G., Conesa, D. Modeling temporal and spatial colony-site dynamics in a long-lived seabird. Popul. Ecol. 45, 133–139 (2003).

[4] Fernández-Chacón, A. *et al.* When to stay, when to disperse and where to go: survival and dispersal patterns in a spatially structured seabird population. Ecography 36, 1117–1126 (2013).

[5] Payo-Payo, A. *et al.* Colonisation in social species: the importance of breeding experience for dispersal in overcoming information barriers. Sci. Rep. 7, 42866 (2017).

[6] Almaraz, P., Oro, D. Size-mediated non-trophic interactions and stochastic predation drive assembly and dynamics in a seabird community. Ecology 92, 1948–1958 (2011).

[7] Oro, D., Pradel, R., Lebreton, J.-D. Food Availability and Nest Predation Influence Life History Traits in Audouin's Gull, Larus audouinii. Oecologia 118, 438–445 (1999).

[8] Payo-Payo, A., Genovart, M., Bertolero, A., Pradel, R., Oro, D. Consecutive cohort effects driven by density-dependence and climate influence early-life survival in a long-lived bird. Proc R Soc B 283, 20153042 (2016).

[9] Tavecchia, G., Pradel, R., Genovart, M., Oro, D. Density-dependent parameters and demographic equilibrium in open population. Oikos 116, 1481–1492 (2007).

[10] Genovart, M., Oro, D., Tenan, S. Immature survival, fertility, and density dependence drive global population dynamics in a long-lived species. Ecology 99, 2823–2832 (2018).

[11] Payo-Payo, A. *et al.* Predator arrival elicits differential dispersal, change in age structure and reproductive performance in a prey population. Sci. Rep. 8, 1971 (2018).

[12] Oro, D., Ruxton, G. D. The formation and growth of seabird colonies: Audouin's gull as a case study. J. Anim. Ecol. 70, 527–535 (2001).

[13] Oro, D. Living in a ghetto within a local population: An empirical example of an ideal despotic distribution. Ecology 89, 838–846 (2008).

[14] Genovart, M. *et al.* Varying demographic impacts of different fisheries on three Mediterranean seabird species. Glob. Change Biol. 23, 3012–3029 (2017).

[15] Genovart, M., Doak, D., Igual, J.M., Sponza, S., Kralj, J., Oro, D. Varying demographic impacts of different fisheries on three Mediterranean seabird species. Glob. Change Biol. 23: 3012–3029 (2014).

[16] J. Frost, Regression Analysis: An Intuitive Guide for Using and Interpreting Linear Models. 1st Edition Statistics by Jim publishing. State College, Pennsylvania State USA (2019)

[17] J. Holland, Hidden Order: How Adaptation Builds Complexity, Helix Books (1996)