Introduction to data fitting methods

Lluís Alsedà

Centre de Recerca Matemàtica

Departament de Matemàtiques
Universitat Autònoma de Barcelona

http://www.mat.uab.cat/∼alseda

October 25 and November 8, 2023

BGSMath Course:

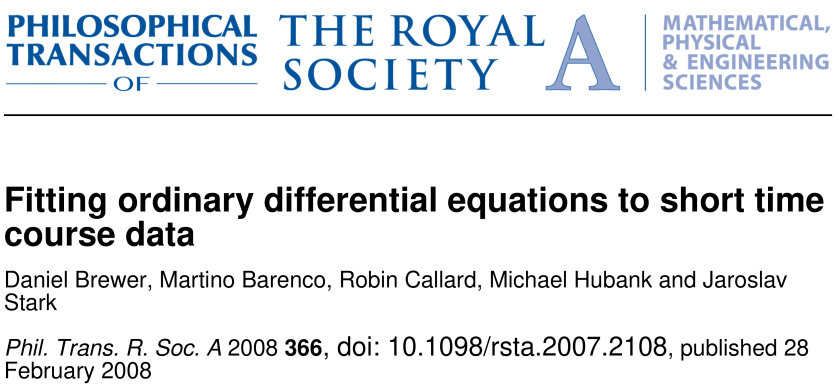
Fitting data with dynamical models:
ten lessons on mathematical field work

Outline

This session will have two main parts:

- An Introduction to data fitting methods
- An application example to the data from the paper *Social copying drives a tipping point for non-linear population collapse*.

For the first part we will heavily (and rather literally) use the paper



To avoid plagiarism, almost literal pieces of this paper will be written in italics, in quotation mode and different colour (as this text).

Introduction

Ordinary differential equations (ODEs) are one of the most popular frameworks for describing the temporal evolution of a wide variety of "real life". Such models take the form

$$\frac{\mathrm{d}\, x(t)}{\mathrm{d}\, t} = f(x(t), t, \alpha)$$

where $x(t)$ is a vector of variables evolving with time; f is a vector field; and α denotes an (in principle optional) set of parameters.

Depending on the model, some parameters can be set empirically, derived from first principles, or measured directly. However, in real-life problems, it is often usual, that many parameters cannot be determined by either of these approaches.

Introduction (continued)

In such situations, one can attempt to estimate the unknown parameters from experimentally measured data. In most cases, these data consist of time series, or time courses, of repeated measurements of one or more experimental variables.

One can fit such data by systematically varying the parameters to determine a set of parameters which minimize the difference between a solution of the differential equation and the data.

A variety of approaches exists to fit data to ODE models. Unfortunately, many of these are poorly documented in the literature, and may only be described in the context of specific applications in specialist publications.

Introduction (continued)

Such methods can be classified in a distinction similar to that between shooting and collocation methods for boundary value problems as follows:

- *The ODE is solved using a conventional iterative numerical integrator (such Runge–Kutta)*
- *or whether a global solution is approximated using splines or related methods.*

Although global methods are now generally preferred for solving boundary value problems, they are poorly developed in the context of fitting ODEs to data. In such cases, shooting is generally more popular and well known.

The hour of truth

Any method for estimating model parameters from data requires two main ingredients.

1. *An error function $E_D(\alpha)$ that quantifies the difference between a model with parameters α and the (observed) data, and*
2. *an optimization method that finds the value of α that minimizes $E_D(\alpha)$.*

The hour of truth: minimizing the error

Except for error functions that have particular features (such as those occurring in linear least squares problems), the minimization stage requires an iterative approach.

Typically, the minimization method can be chosen independently of the construction of the error function.

In some cases, however, an integrated approach can have advantages.

The hour of truth: minimizing the error

A wide variety of standard minimization algorithms exist:

- *When $E_D(\alpha)$ has no, or only a few, local minima apart from the global minimum, then methods that iteratively step downhill, such as the Nelder–Mead simplex method or the Levenberg–Marquardt method work well.*
- *If, on the other hand, $E_D(\alpha)$ has a more complex landscape, stochastic search algorithms such as simulated annealing, Markov Chain Monte Carlo or genetic algorithms are often necessary. These are usually computationally very demanding.*

The hour of truth: minimizing the error

Attempting to estimate the parameters a of the ODE from observed data presents additional difficulties. These are mainly centred on the construction and efficient computation of a suitable error function. This is because we cannot directly determine how well a given set of data points

$$\mathcal{D} := \left\{ \widehat{x}(t_i) : i = 1, 2, \ldots, n \right\}$$

fits the ODE.

Common strategies for the construction and minimization of error functions: Solution-based approaches

By far the better known, and arguably statistically more valid, approach is to solve numerically the EDO to obtain an approximate solution $u(t)$.

Observe that this is meant to provide a model for the data.

Hence, the points $\widehat{x}(t_i)$ should be close to $u(t_i)$.

It is thus natural to base the error function $E_D(\alpha)$ on the difference between $\widehat{x}(t_i)$ and $u(t_i)$. The most common choice is to use the least squares error

$$E_D(\alpha) = \sum_{i=1}^{n} \|\widehat{x}(t_i) - u(t_i)\|^2$$

possibly weighted by the reciprocal of the noise level at each data point.

Common strategies for the construction and minimization of error functions: Solution-based approaches

It is also possible to use this approach even if we cannot measure all of the components of the state vector $\widehat{x}(t_i)$. In such a case, the norm is just taken over the measured components. As long as we have a sufficient number of data points to ensure that $E_D(\alpha)$ has a non-degenerate minimum, it may still be possible to estimate parameters successfully.

When measurement errors are independently normally distributed, $E_D(\alpha)$ will be the logarithm of the likelihood of the data (or more strictly, log (likelihood)), and minimizing equation $E_D(\alpha)$ is equivalent to maximum likelihood estimation of the parameters.

Common strategies for the construction and minimization of error functions: Solution-based approaches

Observe that when f has linear dependence on its parameters, the above error function given is a quadratic form. The minimum in such a case can be obtained efficiently by using algebraic methods such as QR decomposition. This is much faster and more robust than the iterative optimization algorithms mentioned above.

For ODEs this possibility seems irrelevant, however, since even if the vector field f depends linearly on the parameters α, its solution will depend non-linearly on α.

Common strategies for the construction and minimization of error functions: Solution-based approaches

The most popular method of solving the differential equation is to use a standard numerical integration scheme such as Runge–Kutta, possibly with adaptive step-size control. Such an approach immediately encounters a major potential obstacle: we very rarely know the correct initial condition $x(t_0)$ or we have very bad approximations to it (specially when the EDO is highly non-linear).

Common strategies for the construction and minimization of error functions: Shooting

Playing with the initial condition

A better approach is to regard the initial condition as an additional set of unknown parameters which are incorporated in the minimization scheme. We thus regard the error function $E_D(\alpha, x(t_0))$ as depending on both α and $x(t_0)$.

This type of approach appears to have first been tried by Bellman et al. (1967) and similar ideas appear in Swartz & Bremermann (1975). It is closely related to shooting methods for boundary value problems, including methods used for finding periodic orbits and other special solutions (e.g. Golub & Ortega 1992; Kuznetsov 1995). It can work well if data are plentiful and noise levels are low.

Common strategies for the construction and minimization of error functions: Shooting

Playing with the initial condition

However, if we only have a few time points, $E_D(\alpha, x(t_0))$ can have a large number of local minima separated by steep peaks and ridges.

As we shall see below, it is difficult to find the global minimum in such cases. One possible extension is to use multiple shooting methods where the solution is broken down into a number of successive segments, with appropriate matching at the joints (e.g. Kuznetsov 1995; Timmer et al. 2000). This can improve results, but our experience with even moderately complex models is that it can suffer from similar problems to simple shooting.

Other common strategies for the construction and minimization of error functions

- **Collocation methods** *An alternative to using iterative numerical integration is to represent the solution globally using a set of convenient basis functions such as piecewise polynomials (usually of low order); that is splines.*

A list of possible mixed methods is:

1. **Nested minimization and collocation**
   This is closely related to shooting, except that instead of integrating the differential equation using a method such as Runge–Kutta, we use the Newton method to the splines solution at each candidate value of the parameters. This Newton solver is then nested inside an optimization method that iteratively minimizes $E_D(\alpha)$. This is potentially very slow, and we are not aware of this method appearing in the published literature.

Other common strategies for the construction and minimization of error functions

2. **Simultaneous minimization and collocation**
   It is possible to construct combinations of gradient-based minimization and Newton root solving which essentially simultaneously finds the splines solution and minimizes $E_D(\alpha)$.
3. **Dual minimization**
   This consists in dividing the error function $E_D(\alpha)$ into two terms, and minimizing them simultaneously.

- **Derivative approximation methods**
  All of the methods presented so far use the least squares error function. An alternative is to minimize the discrepancy between the r.h.s. and l.h.s. of the differential equation at a selected number of data points.

A comment on the linearity in parameters

In practice, the minimization of the error function can be time-consuming. One would thus like to use specific features of a particular class of models to develop a more efficient algorithms. In particular, many models in physics, chemistry, engineering and biology are linear in their parameters.

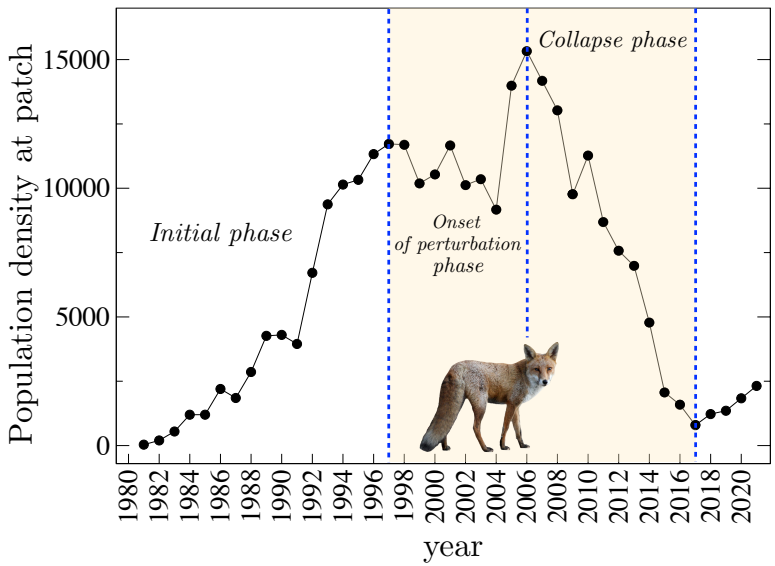
Recall that if a model is linear in its parameters, then the objective function is a quadratic form and a least squares estimate can be obtained in one step (i.e. non-iteratively) using standard linear algebra techniques such as QR decomposition. This is much faster than the iterative minimization routines required for models that are non-linear in parameters.

A comment on the linearity in parameters

It seems difficult, however, to make use of linearity with respect to parameter's for ODE models, since even if the vector field depends linearly on the parameters, the solution will typically exhibit non-linear dependence.

Part II: The data

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		



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. 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).

The Model

$$\frac{\mathrm{d}}{\mathrm{d}t}x(t) = (\vartheta+\omega)x(t)\left(1-\frac{x(t)}{K}\right) - \varepsilon x(t) - \Big[\rho x(t) + \lambda\Psi(x(t))\Big]$$

$$= \boxed{\varphi x(t)} - \boxed{\beta x(t)^2} - \boxed{\lambda\Psi(x(t),\mu,\sigma,\delta)}$$

Immigration, growth and death

Nonlinear competition term

Dispersal by social copying

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
ρ	year^{-1}	$\mathbb{R}^+$	Linear (exponential) dispersal rate
λ	year^{-1}	$\mathbb{R}^+$	Dispersal rate by social copying

Dynamics after the perturbation.
The migration term: dispersal by social copying

$$\Psi(x,\mu,\sigma,\delta) := \begin{cases} \frac{1-\mathcal{E}_{\text{dir}}(x,\mu,\sigma,\delta)}{1-\mathcal{E}_{\text{dir}}(0,\mu,\sigma,\delta)} & \text{when } 0 \leq x \leq \delta, \\ \frac{1-\mathcal{E}(x,\sigma,\delta)}{1-\mathcal{E}_{\text{dir}}(0,\mu,\sigma,\delta)} & \text{when } x \geq \delta, \end{cases}$$

where

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

and

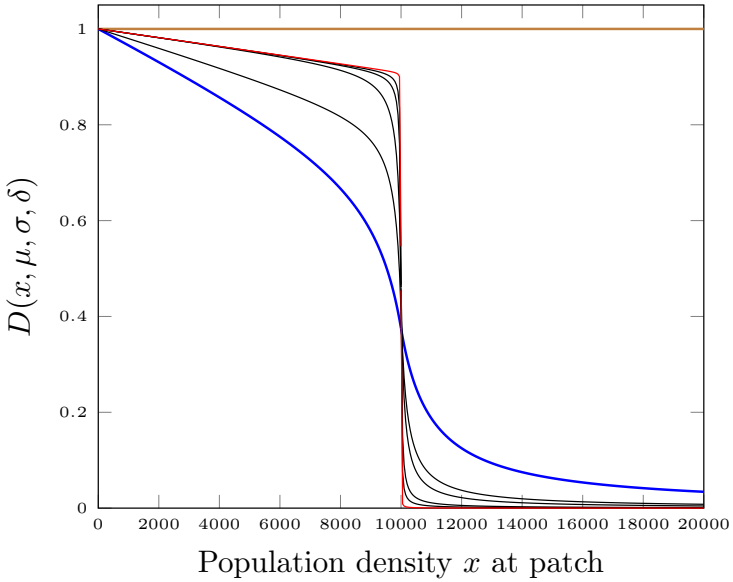
$$\mathcal{E}(x,\sigma,\delta) := \frac{\sigma(x-\delta)}{\Theta+\sigma|x-\delta|},$$

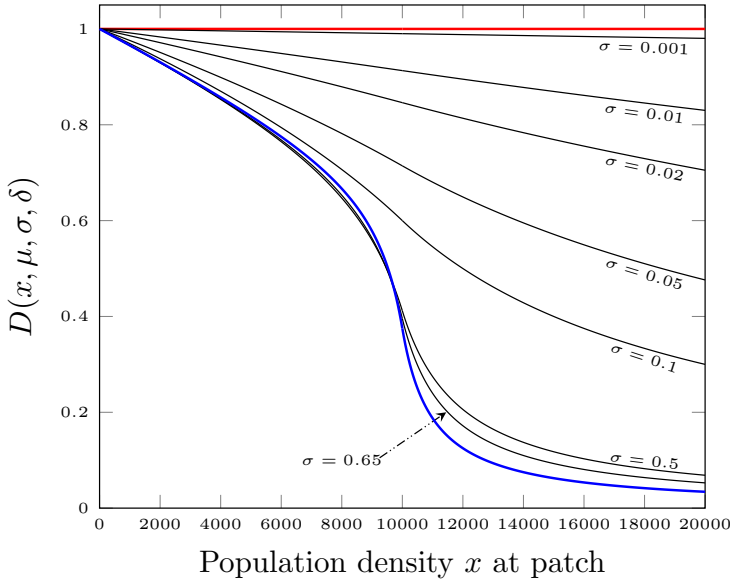
is an *Elliot sigmoid* Θ–scaled, σ–strengthened, and δ–displaced. All the parameters of the dispersal function are non-negative and we have fixed $\Theta := 1000$ (this parameter controls the scale in the independent variable x which is related with the order of magnitude of the carrying capacity K).

The parameters

Parameter	Units	Range	Ecological meaning or description
$\alpha = \vartheta + \omega - \varepsilon$	year^{-1}	$(-\infty, \vartheta + \omega]$	Population growth rate including death of individuals (without linear dispersal)
$K = \frac{\alpha+\varepsilon}{\beta}$	birds	$[1, +\infty)$	Carrying capacity
$x(0)$	birds	$[0, K]$	Initial condition
$\varphi = \alpha - \rho$	year^{-1}	$(-\infty, \alpha]$	Neat population growth rate including linear dispersal
β	(birds $\times$ year)$^{-1}$	$[0, +\infty)$	Intrinsic growth rate over the carrying capacity
Parameters concerning dispersal rate by social copying			
λ	year^{-1}	$\mathbb{R}^+$	Dispersal rate by social copying
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

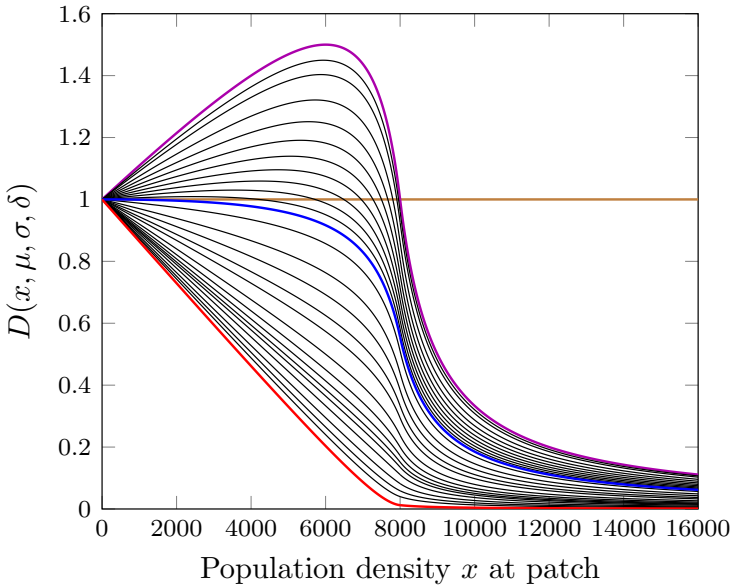
The role of σ in $\Psi(x, \mu, \sigma, \delta)$

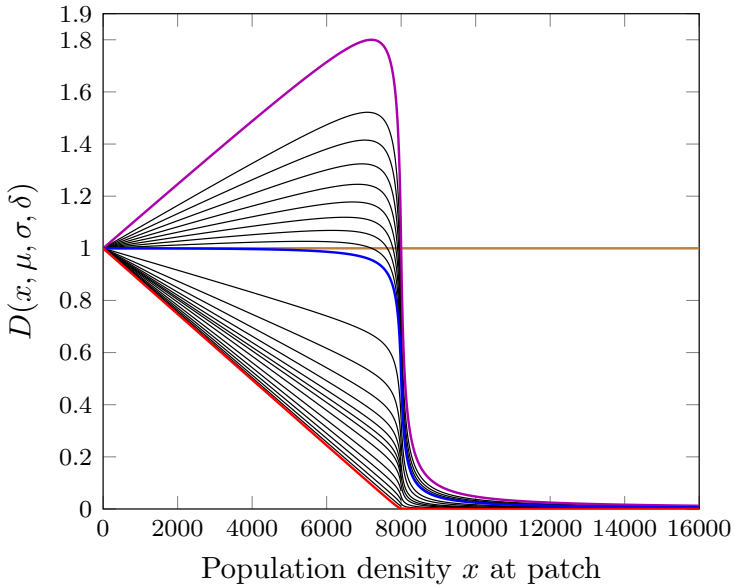




Shapes of the function $\Psi(x, \mu, \sigma, \delta)$ used to model social copying behaviour during dispersal. We explore the ranges of $\sigma \geq 1$ (left panel) and $\sigma < 1$ (right panel) fixing $\delta = 10^4$.
Left: The brown graph corresponds to $\sigma = 0$, involving density-independent 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 $\sigma = 0.65, 0.5, 0.1, 0.05, 0.02, 10^{-2}, 10{-}3$.

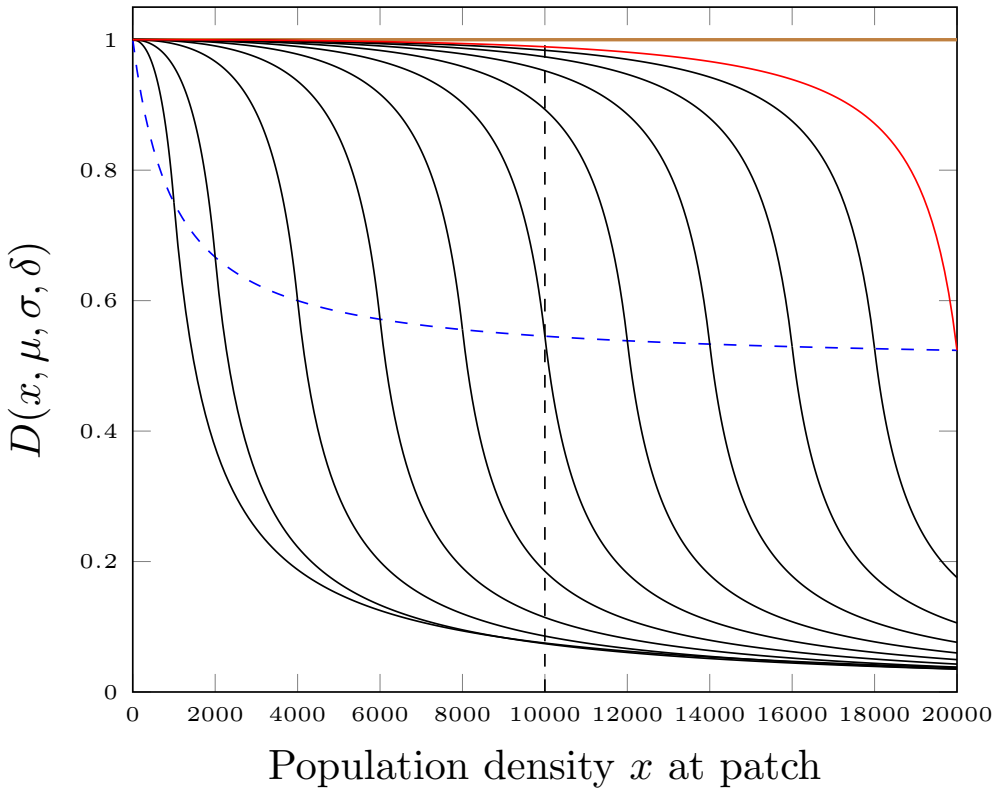
The role of μ in $\Psi(x, \mu, \sigma, \delta)$





Shapes of the dispersal function $\Psi(x, \mu, \sigma, \delta)$ 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 $\mu > 1$.

The role of δ in $\Psi(x, \mu, \sigma, \delta)$



Shapes of the dispersal function $\Psi(x, \mu, \sigma, \delta)$ 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 $\delta = 10^4$.

Model fitting and parameters estimation: Collapse phase 2006–2017

The solution of the model with β fixed to $2.4382635446 \times 10^{-5}$, and its parameters belonging to the ranges shown in the above table 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.

The observed population of Audouin's gulls at the years 2006 to 2017 is denoted by

$$\pi(\ell, \ell = 0:11) = \text{Audouin's\_Gulls\_Observed\_Population\_at\_year}(2006 + \ell, \ell = 0:11) =$$

$$\begin{matrix} [0] & [1] & [2] & [3] & [4] & [5] & [6] & [7] & [8] & [9] & [10] & [11] \end{matrix}$$
$$\big[15329, 14177, 13031, 9762, 11271, 8688, 7571, 6983, 4778, 2067, 1586, 793\big].$$

Parameter space and Fitting Function

$$\mathscr{P} := [0, K] \times [-\infty, \alpha] \times \mathbb{R}^+ \times \mathbb{R}^+ \times \mathbb{R}^+ \times \mathbb{R}^+,$$

$$\begin{aligned} \mathrm{F}: \quad \mathscr{P} &\longrightarrow \mathbb{R}^+ \\ (x(0), \varphi, \lambda, \mu, \sigma, \delta) &\longmapsto \sqrt{\textstyle\sum_{\ell=0}^{11} (x(\ell) - \pi(\ell))^2}. \end{aligned}$$

The *fitting of the model*

consists in solving

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

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

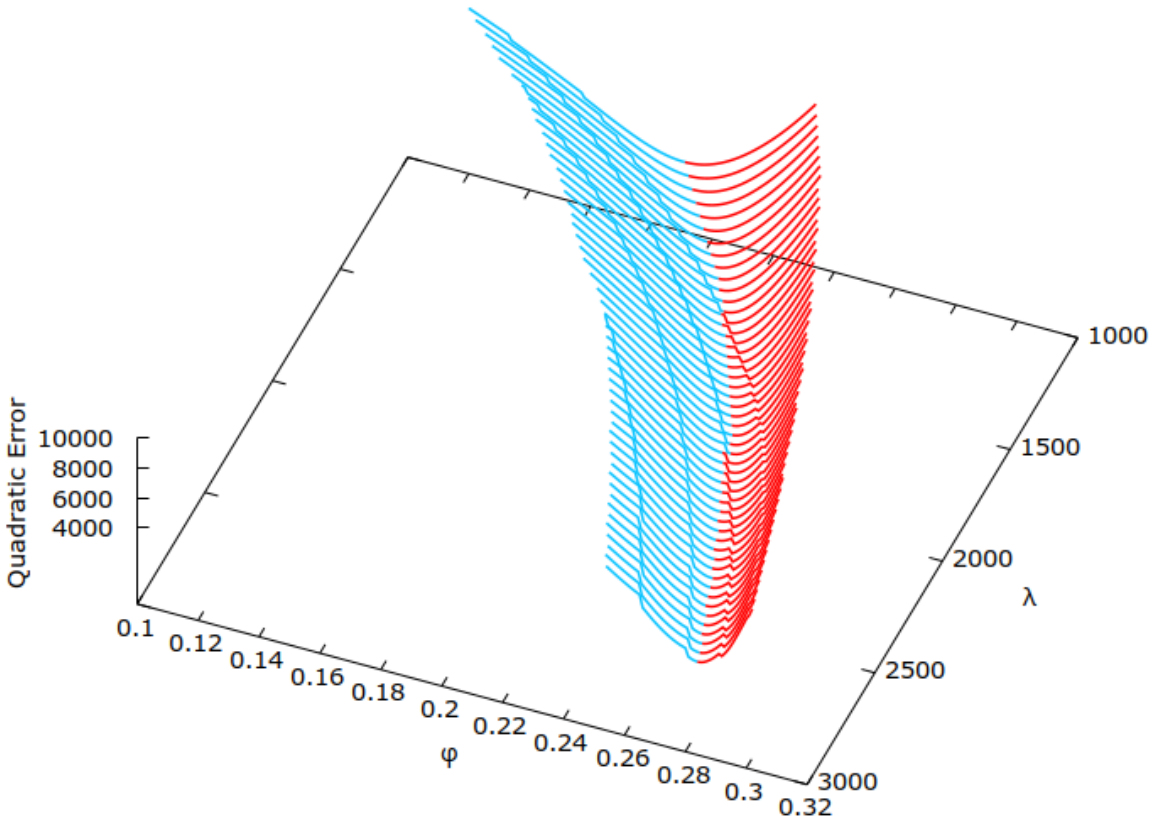
Observe that

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

has 0 as a lower bound.

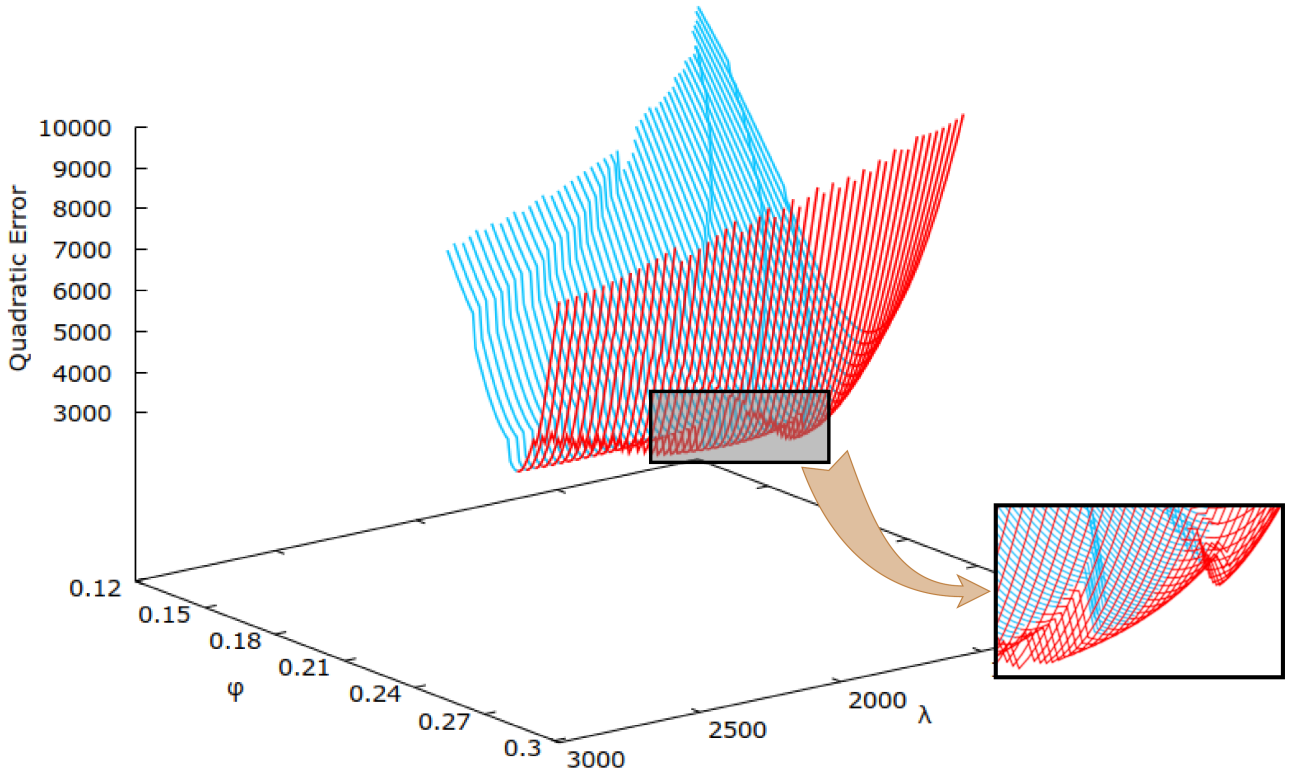
Since the map F is continuous on $\mathscr{P}$ we could guarantee that the above minimization problem has at least one solution *provided that the parameter space $\mathscr{P}$ is compact.*

Part of the landscape



A view of the landscape of the function F around the point $(15800, 0.22, 1400, 0, 1, 8740)$ showing only points whose 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 plot in the next slide shows a zoom of the "valley" of the landscape from the upper picture where a complicate local minima distribution is seen.


Quadratic Error
10000
9000
8000
7000
6000
5000
4000
3000
0.12
0.15
0.18
0.21
0.24
0.27
0.3
φ
3000
2500
2000
λ

On the reduction to a compact parameter space $\mathscr{K}$

We compulsory need a relatively small compact set $\mathscr{K} \subset \mathscr{P}$ that contains the minimum (or equivalently that $\mathscr{P} \setminus \mathscr{K}$ does not contain the minimum) of the 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 the above minimization problem.
- Second, the reduction of the parameters' search space from $\mathscr{P}$ to $\mathscr{K}$ will make possible the use of standard minimization algorithms.

On the reduction to a compact parameter space $\mathscr{K}$ (II)

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 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 on $\mathscr{P} \setminus \mathscr{K}$, and to find a point from $\mathscr{P}$ reasonably close to the optimum.

The *Sparse Anisotropic Grid Search (SAGS)*

$$\mathscr{S} := [12726, 17932] \times [0.12, \alpha] \times [300, +\infty] \times \mathbb{R}^+ \times \mathbb{R}^+ \times \mathbb{R}^+ \subset \mathscr{P}$$

$$\mathscr{L} := [12726, 17932] \times [0.12, \alpha] \times [300, 100000] \times [0, 600] \times [0, 5000] \times [0, 159000] \subset \mathscr{S}, \text{ and}$$

$$\mathscr{K} := [12726, 17932] \times [0.12, \alpha] \times [300, 3000] \times [0, 10] \times [0, 50] \times [0, 20000] \subset \mathscr{L}.$$

		Sparse Anisotropic Grid Search		Large Domain Grid Search (SALDGS)	
Parameter	**Theoretical Range**	**Effective Search Range**	**Anisotropic Step**	**Effective Search Range**	**Anisotropic Step**
$x(0)$	$[0, K]$	$[12600, 18800]$	200	$\mathscr{K} \setminus \mathscr{L}$	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

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 $F(x(0), \varphi, \lambda, \mu, \sigma, \delta)$ 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 $\mathscr{L}$.

Conclusions of the Sparse Anisotropic Grid Search

- $\mathrm{F}\big(15800, 0.22, 1400, 0, 1, 8740\big) = 2602.4358676183260\cdots$ and $x_{(15800,0.22,1400,0,1,8740)}(t) \in [0, K]$ for $t \in [0, 11]$.
- For every point $\vec{\boldsymbol{\theta}} \in \mathscr{L} \setminus \mathscr{K}$ whose components belong to the grid described in the table above we have

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

- A Montecarlo exploration on the computer-representable part of the region $\mathscr{S} \setminus \mathscr{L}$, gives

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

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

A useful basic argument towards finding an appropriate compact domain

Lemma

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

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

Consequently,

$$\arg\min_{\vec{\boldsymbol{\theta}} \in \mathscr{P}} \mathrm{F}\big(\vec{\boldsymbol{\theta}}\big) \in \mathscr{S}.$$

A useful basic argument towards finding an appropriate compact domain: The details

To prove the previous lemma we use the following analytical result.

Bounding Lemma

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 $t \in [0, T^*]$.

A useful basic argument towards finding an appropriate compact domain

The trivial proof of $x(0) \in (12726, 17932)$

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 the previous lemma we have,

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

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

Analogously, if $\kappa \geq 17932$,

$$\mathrm{F}\big(\kappa, \varphi, \nu, \lambda, \mu, \sigma, \delta\big) \geq \kappa - \pi(0) \geq$$

$$17932 - 15329 > \mathrm{F}\big(15800, 0.22, 1400, 0, 1, 8740\big).$$

A useful basic argument towards finding an appropriate compact domain

The not so trivial proof that $\varphi > 0.12$

Now assume that $\varphi \leq 0.12$. We denote by $u(t)$, $t \in [0, 11]$, the solution of the model $\dot{u} = \widetilde{\varphi} u - \beta u^2$ with $\widetilde{\varphi} = 0.12$, $\beta = 2.43826356697 \times 10^{-5}$, and initial condition $u(0) = 17932$. This is a Ricatti Equation that has analytical solution:

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

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

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

and

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

Since the function $\Psi(x, \mu, \sigma, \delta)$, as a function of x, is continuous, differentiable, and strictly positive,

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

for every $x \in \mathbb{R}^+$. Then, since $x(0) \leq 17932 = u(0)$, by the Bounding Lemma(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 $\ell = 1, 2, 4, 5$.

A useful basic argument towards finding an appropriate compact domain

The not so trivial proof that $\varphi > 0.12$

Hence,

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

$$\sqrt{\sum_{\ell\in\{1,2,4,5\}}\big(\pi(\ell)-x(\ell)\big)^2} \geq \sqrt{\sum_{\ell\in\{1,2,4,5\}}\big(\pi(\ell)-u(\ell)\big)^2} >$$

$$2885 > \mathsf{F}\big(15800, 0.22, 1400, 0, 1, 8740\big).$$

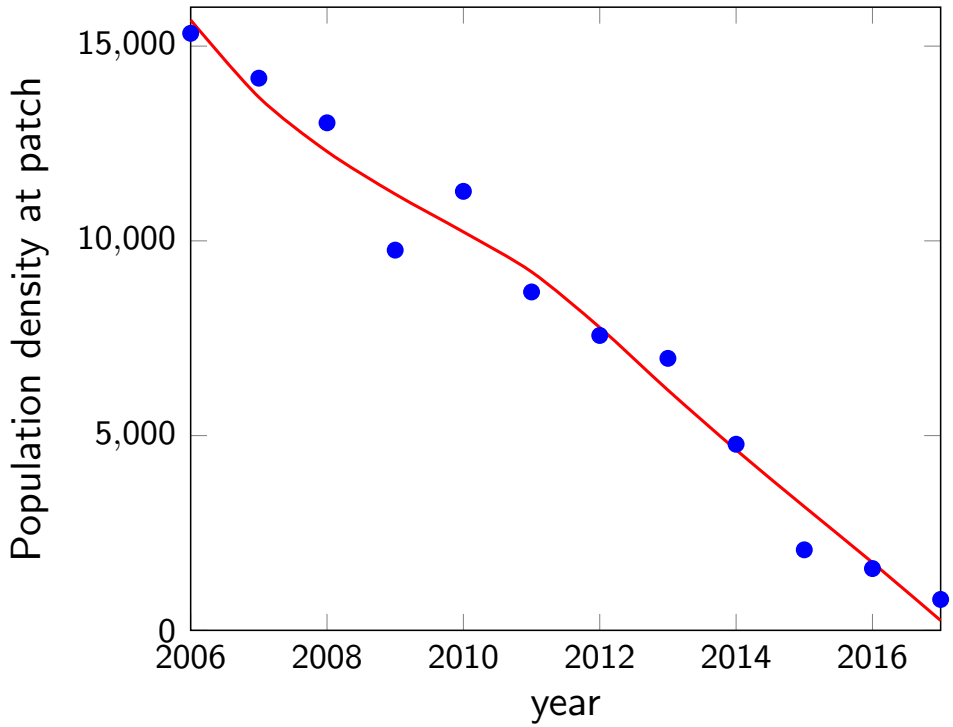
The solution and the modelling conclusions

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	0.490475636435769 0
δ	8944.2282749675759987
Quadratic Error = $F(x(0), \varphi, \lambda, \mu, \sigma, \delta)$	2566.999667640135158

The solution and the modelling conclusions

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



The solution and the modelling conclusions