

OPEN

Meta-modeling on detailed geography for accurate prediction of invasive alien species dispersal

Nick Pepper^{1*}, Luca Gerardo-Giorda^{2,3} & Francesco Montomoli^{1,3}

Invasive species are recognized as a significant threat to biodiversity. The mathematical modeling of their spatio-temporal dynamics can provide significant help to environmental managers in devising suitable control strategies. Several mathematical approaches have been proposed in recent decades to efficiently model the dispersal of invasive species. Relying on the assumption that the dispersal of an individual is random, but the density of individuals at the scale of the population can be considered smooth, reaction-diffusion models are a good trade-off between model complexity and flexibility for use in different situations. In this paper we present a continuous reaction-diffusion model coupled with arbitrary Polynomial Chaos (aPC) to assess the impact of uncertainties in the model parameters. We show how the finite elements framework is well-suited to handle important landscape heterogeneities as elevation and the complex geometries associated with the boundaries of an actual geographical region. We demonstrate the main capabilities of the proposed coupled model by assessing the uncertainties in the invasion of an alien species invading the Basque Country region in Northern Spain.

The spread of invasive species is responsible for severe ecological damage and economic losses and is recognised as a significant threat to biodiversity^{1,2}. Its management is a challenging problem due to the uncertainty surrounding the species characteristics and the usually limited resources allocated to control. Firstly, measures to protect the native species of a region from invasive species are costly for the local government. For example, the Basque water agency (URA) budgets an average of €1.4 million a year for activities related to the elimination of invasive species and the restoration of habitats. In addition to this figure, the Biodiversity Strategy of the Basque Autonomous Community 2030 calls for €1,119,470 to be spent on the protection and restoration of habitats in 2019 alone³. Moreover, the management often starts long after the appearance of the invasive species, when the ecological damage is already visible and the species well-established, taking an additional financial toll. Control strategies may aim at extirpation or rely on forcing a component Allee effect by bringing the population below a threshold⁴, for instance by increasing the abundance of predators, by culling or inducing habitat fragmentation⁵.

Another important consideration is the need to account for the temporal dynamics of the invader population. Managers have to optimize their choices not only for the present, but must also take into account possible future scenarios. The introduction of *cane toad* to Australia from Hawaii in 1935 in an attempt to control the native grey-backed *cane beetle* and its subsequent, uncontrolled, spread throughout the continent is a cautionary tale in this respect⁶. On the other hand, if the understanding of population persistence and competitive outcomes can provide long-term information, then short-term predictions of spread are often more important in the management of invasive species as fast responses can potentially save public administrations sizeable amounts of money⁷. Optimal management strategies generally rely on the coupling of biological models with optimization procedures and are reviewed in Epanchin-Niell and Hastings⁸. Within such a framework, the modeling of invasive species and the dynamics of their dispersal can be of great help to environment conservation agents and policy makers.

A number of works have appeared in recent decades which have aimed to mathematically model invasive species at different levels of accuracy. A popular technique in devising predictive models has been Spatially Explicit Population dynamics Models (SEPMs). Introduced by Rushton and collaborators in 1997⁹ in the study of red and grey squirrel distribution at landscape scale, SEPMs are models that combine GIS data and a population dynamics approach on separated habitat blocks. These models have been used quite effectively to study the invasive expansion¹⁰ and control¹¹ of the grey squirrel in northern Italy. The major drawback of SEPMs is that their performance

¹UQLab, Department of Aeronautics, Imperial College London, SW7 2AZ, London, UK. ²BCAM – Basque Center for Applied Mathematics, Bilbao, Spain. ³These authors contributed equally: Luca Gerardo-Giorda and Francesco Montomoli. *email: nicholas.pepper16@imperial.ac.uk

is sensitive to the accuracy of the estimates of the life-history parameter input. As a consequence, SEPMs can only be used for species for which detailed data of population parameters (fecundity, mortality, dispersal distance, density) in different types of habitat are available. Expanding on the ideas of SEPMs, Jones *et al.*¹² recently used a Gis-Grass based deterministic model with stochastic rates¹³ to study control strategies to protect native red squirrels from invasive grey squirrels on the island of Anglesey in the UK. Catford *et al.*¹⁴ used boosted regression trees to identify areas within a region of interest that are the most susceptible to invasion by non-native plant species. The model is not species specific and uses a set of 18 uncorrelated variables to classify the most vulnerable regions, which can be prioritized in conservation efforts. Gassó *et al.*¹⁵ studied the distributions of native and invasive plant species in Catalonia, Spain and found that human influence on the landscape had a greater influence on the species richness of the invasive species than climatic variables. Campos *et al.*¹⁶ used a statistical approach on patches to investigate the effect of environmental factors on the invasion patterns of 89 invasive plants species in the Basque Country, concluding that the spatial structure of alien plant invasions in the region are strongly environmentally induced. A common trait of these works is that the impact of geography on the distribution of invasive species is recognised, even if the exact nature of the complex link between the landscape and the abundance of an invasive species is not completely understood¹⁷.

Accounting for the heterogeneities of the landscape in invasive species modeling is not a trivial task. Forests, plains and mountains present different levels of hospitability to the invasive species, whilst strong natural barriers such as motorways, lakes and major waterways can reduce, if not stop completely, the spread of the species¹⁸. Different approaches have been used in the literature to cope with the spatial component of heterogeneities, from data-driven statistical modeling to deterministic systems of differential equations. Lohr *et al.*¹⁹ investigated the effect of various land management options on local biodiversity via a system of ordinary differential equations for the abundance of two competing species, one local and one invasive, on an island. Although able to account for several important factors such as growth rate, carrying capacity, species interaction, efficacy and cost of management, their method is spatially lumped on each type of habitat. While on the one hand such a model can be very effective in the management of an already established population, it lacks the spatial accuracy required to predict an invasion and thus is of limited use in devising early interventional strategies.

Meier *et al.*²⁰ combined species distribution models²¹ (SDMs) with population spread models²² (models which cover meta-population dynamics) to evaluate the effectiveness of control actions on three widespread invasive alien plants for different spending scenarios and management goals in a complex landscape on the Swiss plateau. They showed how control actions under a restricted budget are more effective if spatially prioritized and, at the same time, that applying intensive control at early stages generally increased the effectiveness of control. In a recent paper, Maciel and Lutscher²³ studied the coexistence of two competitive species in dependence of habitat preference, diffusion rates and carrying capacity. In order to perform their analysis, based on homogenization techniques and a multi-scale approach, they needed to consider an idealized situation where the two species are very similar and located in a one-dimensional, patchy, infinite landscape where the distribution of habitat patches is periodical and alternates between strongly favourable and mildly favourable to the species. As a consequence, although this approach provides interesting insight on the competing dynamics, it cannot be immediately translated to a realistic landscape.

Since observational data is in general collected according to the geopolitical structure of a region, the intuitive idea behind the works presented above is to subdivide the region of interest into discrete units either geopolitically or territorially. A geopolitical subdivision follows administrative subdivisions at various levels such as at the levels of townships, counties or provinces, whilst a territorial subdivision divides the landscape into regular territorial patches and considers the species movements from one unit to another. Geopolitical decompositions have been used extensively in epidemiology, for example in Smith *et al.*²⁴ where the spatial propagation of raccoon rabies in Connecticut, USA is modelled. Although a priori well adapted to observational data, a geopolitical approach suffers from significant limitations when considering wildlife. On the one hand, animals do not in general move at the geopolitical scale, nor do their movements fit the shape of geopolitical units. On the other hand, plants do not move and their diffusion is driven by the dispersal of seeds that can be transported by climate agents such as wind or rivers, but in general do not move very far from the plant that generated them. As a consequence, several scientists have applied continuous reaction-diffusion models based on Fisher's equation²⁵ to problems in ecology. Reaction-diffusion models rely on the implicit assumption that the dispersal of an individual is random but at the scale of the population the density of individuals is smooth and the invasion front is somewhat regular²⁶. Recently, Bonneau *et al.*²⁷ applied a reaction diffusion model to simulate the effect of management efforts on the distribution of Burmese pythons in the Everglades (Florida, USA). The reaction-diffusion method was shown to yield better prediction accuracy than the more established Constant Effort Harvesting model. A limitation of the study is that the region of interest was treated as an isolated environment with zero flux boundary conditions imposed. This is a common feature to many works (not only in reaction-diffusion models), where the considered environment is either an island or a region bounded by impenetrable borders and where the effect of migration from neighboring domains is not considered^{12,19,28–30}.

Although reaction-diffusion models are a good trade-off between model complexity and flexibility for use in different situations, their limited application resides in the technique used in their numerical approximation. Common reaction-diffusion models are approximated in space by means of finite differences methods (FD, see, e.g.³¹). The natural framework of FD requires the use of a Cartesian discretization grid. As the latter cannot adapt to the complex geometries that represent the boundaries of a real geographical region, researchers using reaction-diffusion models have limited themselves to either idealized geometries (one dimensional intervals²³ or rectangles^{27,32}) or covered the land under study through lattices composed of rectangular patches with sizes that are too large to cover smaller heterogeneities. As an example, Hooten and Wikle³³ used, in a Bayesian setting, a finite difference approach on a lattice to approximate the latent process for the logarithm of the intensity of a Poisson process describing the abundance of the Eurasian Collared dove in the South East of the USA. In his PhD

Thesis, Arab³⁴ suggested an extension to this approach by using finite elements in the approximation of the latent process, but both his derivation and implementation never went past a one-dimensional interval.

In this paper, we consider a continuous, time-dependent, nonlinear reaction-diffusion model for the species density that arises from a generalization of the Fisher equation. Following an idea introduced in Keller *et al.*³⁵, in which the spread of rabies among raccoons in New York State was studied, landscape heterogeneities are accounted for by including in the computational domain the significant geographical features of the area by acting directly on the coefficients of the model that are both spatially and temporally dependent. In particular, we show how the model can account for elevation in a natural way. Elevation is a major trait in landscape heterogeneity, pretty much overlooked in the literature, despite the fact that quite often species' habitats are bounded by it¹⁸. Moreover, recent studies focusing on the impact of climate change in modifying the habitat of species have found that increased temperature and changes in humidity have rendered higher altitudes suddenly hospitable for new species that might pose a threat to indigenous ones (see e.g.^{22,36–39}). From this perspective, the simplicity with which the model proposed here is able to account for elevation can become a major asset. The reaction-diffusion model is numerically approximated by the finite element method, which is able to treat arbitrarily shaped boundaries, like the ones of a geographical region⁴⁰. Moreover, finite elements can easily be employed on adaptive grids that feature finer resolution in specific regions of interest, for instance co-localized with peculiar landscape characteristics.

For a tool aimed at supporting management strategy in environmental conservation it is necessary to provide not only responses on suitable quantities of interest commonly used in the decision process, but also confidence intervals. To cope with how the uncertainty in the data and the characteristics of the species are propagated by the proposed model, we couple the reaction-diffusion equation with arbitrary Polynomial Chaos (aPC, see, e.g.⁴¹). Quantifying uncertainty is important for predictive models, and may become paramount if conservation agencies are faced with the threat from a species about which very little is known. Arbitrary Polynomial Chaos, being very efficient in its handling of scarce data, appears a natural choice.

This paper is organized as follows. In Section 2 we introduce a deterministic continuous advection-diffusion-reaction term model for the dispersal of an invasive species in an heterogeneous environment; describe the way it accounts for geographical characteristics and present the simulation protocol. In Section 3 the deterministic model is demonstrated by its application to an alien species invading the Basque Country region in Northern Spain. Section 4 introduces the arbitrary Polynomial Chaos (aPC) method used to propagate parametric uncertainties through the model. The method is then applied to assess uncertainty on the test case introduced in Section 3 using synthetic data.

Methods

A general distributed model for species dynamics. The spread of an invasive species in a region denoted by $\Omega \subset \mathbb{R}^2$ can be modeled by a modified Fisher's equation for the density $u(\vec{x}, t)$ of individuals in location \vec{x} at time t (see⁴² for an introduction to diffusion in ecological problems). In the more general form the model is a nonlinear advection-diffusion-reaction equation of parabolic type:

$$\partial_t u - \operatorname{div}(\nu(\vec{x}, t)\nabla u) + \mathbf{b}(\vec{x}, t) \cdot \nabla u + u(u - \gamma(\vec{x}, t)) = f(\vec{x}, t) \text{ in } \Omega \times [0, T]. \quad (1)$$

In Eq. (1) the diffusion coefficient $\nu(\vec{x}, t)$ accounts for the static landscape heterogeneities such as inhospitable regions and the presence of natural or artificial barriers within the computational domain. Examples of such barriers include major waterways and roads. The carrying capacity $\gamma(\vec{x}, t)$ expresses the birth and death rates of the species at a given spatial and temporal location. Their spatial dependence allows the inclusion of peculiar landscape characteristics in terms of the classification of the terrain (town, meadow, rock, wood, elevation). The convective term $\mathbf{b}(\vec{x}, t) \cdot \nabla u$ allows for transport effects within the region to be modeled. In the case of terrestrial amphibious animals, the transport effects are associated with the presence of rivers that can act both as an accelerator or as a contrasting agent according to their flow direction with respect to the direction of the propagation front; outside of waterways the transport effects vanish. The temporal dependence of the coefficients allows for various types of seasonality: variations in the level of water and flow velocity of rivers, breeding for animals or sprouting for plants. Considering seasonality can be very important since, as shown by Meier *et al.*²⁰, the most effective spatial treatments against invasive species prioritize small populations in the case of the seasonal species and large populations in the case of the perennial species.

Equation (1) is completed by a suitable initial value $u^0(\vec{x})$ and boundary conditions on $\partial\Omega$ ⁴³. In general, homogeneous Neumann boundary conditions are used to model an isolated environment, but other boundary conditions can be considered as well: a homogeneous Dirichlet boundary condition would model a hostile environment, while a Robin boundary condition would model migratory dynamics^{27,42}. Finally, let us point out that long distance dispersal could be included in the model by adding a stochastic term to the reaction-diffusion equation, but this goes beyond the scope of this paper.

Geographical accuracy of the model. *Elevation.* A geographically detailed description of the region of interest should not only include evident landscape heterogeneities such as rivers, lakes or urban areas, but also elevation. The latter can play a crucial role in characterising the hospitality of an area for a given species. Commonly both animal and plant species are suited to habitats in a bounded region of elevation and this aspect should be taken into account. The NASA Shuttle Radar Topography Mission (SRTM, February 11 to 22, 2000) produced a near-global covering of land on Earth, although limited to latitudes between 60S and 60N. Files containing the elevation data have been made publicly available (the elevation data has been collected in files available at http://dds.cr.usgs.gov/srtm/version2_1). Elevation values are integers (INT16 class): sea level values are 0, unknown values equal -32768 (there is no NaN for INT class). Each file corresponds to a tile of 1×1 degree of a

square 1201×1201 grid, consisting of latitude, longitude, and elevation values (SRTM3 = 3 arc-seconds). For territories within the USA the tiles feature a higher resolution of 3601×3601 grid (SRTM1 = 1 arc-second). By using this data, a three dimensional grid may be generated for a given region of interest that lies between 60S and 60N. In what follows, we will denote by $z = z(\vec{x})$ the elevation at point $\vec{x} \in \Omega$.

Extended heterogeneities and carrying capacity. Strong, extended heterogeneities may significantly affect the dynamics of the dispersal across a whole region. To cope with such features we take advantage of the temporal and spatial dependence of the carrying capacity and the diffusion coefficient in Eq. (1). In general, we consider the carrying capacity to be naturally dependent upon the elevation of the ground, however, different levels of hospitality can occur at a given elevation due to the landscape heterogeneities and such features can be taken into account. Less hospitable regions are modeled by lowering the carrying capacity to reduce the population density at equilibrium (see also²³). The diffusion coefficient is also reduced to minimize the dispersal towards such inhospitable regions. Lakes are included in the model by setting the carrying capacity to 0 in their interior.

Treatment of boundary condition artifacts. Geographical regions are usually characterized by fairly irregular boundaries. Few approaches in the literature use real geographical boundaries or impose no-flux boundary conditions in the cases that do, modeling the region of interest as an isolated environment. Whilst this may be a reasonable assumption if the region of interest is an island, in the case of a territory continuous with surrounding areas the choice of restricting the modeling to the actual region of interest could trigger unrealistic features in the vicinity of its boundary. To avoid such an effect, a computational domain larger than the region of interest is considered. This approach is commonly used in Computational Fluid Dynamics (CFD) simulations in aeronautics⁴⁴: if the computational domain is sufficiently large, a zero flux boundary condition can be imposed without impairing the simulation in the region of interest.

Numerical approximation. The finite element method (FEM, see⁴⁰ for an introduction to the subject) is particularly suitable for complex geometries like the ones represented by real geographical regions. We thus discretize Eq. (1) in space by means of finite elements, while a classical finite difference discretization is carried out for the time discretization.

Spatial discretization. For a regular triangulation (also called mesh) of the domain Ω , a finite element space of order k consists of globally continuous functions that are locally a polynomial of degree at most k on every triangle of the mesh. Each function of a finite element space can be represented as a linear combination of a suitable finite basis. In the numerical simulations presented in this paper we will use first order elements. Although higher order polynomials can be considered⁴⁰, linear ones feature a sufficient level of accuracy for the applications being considering here. A generic element of the first order basis, denoted with $\varphi_j(\vec{x})$ ($j = 1, \dots, N$, N being the total number of mesh points), is the piecewise linear function equal to 1 on the j^{th} node of the mesh and equal to 0 in all the other nodes. The finite element approximation of the solution of Eq. (1) is given by

$$\vec{u}_h(\vec{x}, t) = \sum_{j=1}^N u_j(t) \varphi_j(\vec{x}). \quad (2)$$

The unknown time-dependent vector $\vec{u}(t) = [u_1(t), \dots, u_N(t)]^T$ solves the N -dimensional nonlinear ordinary differential system

$$M \frac{d\vec{u}(t)}{dt} + A \vec{u}(t) + B \vec{u}(t) = \mathcal{F}(\vec{u}(t)), \quad (3)$$

where M , A , and B are the mass, stiffness and transport matrices, whose (i, j) -th entries are

$$M_{ij} = \int_{\Omega} \varphi_j \varphi_i d\vec{x}, \quad A_{ij} = \int_{\Omega} (\nu(\vec{x}) \nabla \varphi_j) \nabla \varphi_i d\vec{x}, \quad B_{ij} = \int_{\Omega} (\vec{b}(\vec{x}) \cdot \nabla \varphi_j) \varphi_i d\vec{x}, \quad (4)$$

while $\mathcal{F}(\vec{u}(t)) = \int_{\Omega} (\gamma(\vec{x}, t) - \vec{u}_h(t)) \vec{u}_h \cdot \varphi_i d\vec{x}$ is the discretization of the nonlinear term. All of the above integrals are computed by means of suitable quadrature rules.

Time discretization. Let Δt be a time step and $t^n = n\Delta t$ be the discretization of the time interval under study. We denote by $\mathbf{u}^n = [\mathbf{u}_j(t^n)]_{j=1, \dots, N}$ the vector of the nodal values at time t^n . The incremental ratio $(\mathbf{u}^{n+1} - \mathbf{u}^n)/\Delta t$ is an approximation of the time derivative either in t^{n+1} or t^n , with an associated numerical error proportional to Δt . As a good trade-off between numerical stability and computational efficiency, Eq. (3) is advanced in time by a mixed implicit/explicit (IMEX) approximation scheme, where the stiffness and the transport are treated implicitly while the nonlinear term is treated explicitly. Knowing the approximation \mathbf{u}^n of the solution at time step t^n , the solution at time step t^{n+1} is obtained by solving

$$(M + \Delta t A + \Delta t B) \mathbf{u}^{n+1} = M \mathbf{u}^n + \Delta t F(\mathbf{u}^n). \quad (5)$$

Equation (5) is a linear system of dimension N , where the right hand side is easily built since \mathbf{u}^n is known. For a more extensive discussion and application of the IMEX method see³¹.

Simulation protocol. The numerical simulations of Eq. (1) are performed with a self-developed code in Matlab (MathWorks Inc., Natick, MA) with a uniform time step of $\Delta t = 0.1$ months. Several techniques to efficiently solve linear systems like (5) are available based on iterative methods: at every time step we solve (5) with the conjugate gradient method, preconditioned by an incomplete Cholesky factorization (see, for instance the book by Y. Saad⁴⁵). The Matlab source code is available as supplementary material to this paper.

Data driven mesh generation. The computational domain is discretized by a triangular mesh to exploit its higher accuracy (compared to a typical cartesian grid for finite differences approximations) in representing both complex boundaries and peculiar structures in the interior of the domain. Adaptive grids are used in order to harness the best approximation properties of the finite element method. These feature smaller triangles in the areas that require higher levels of accuracy in the landscape, like the region of interest and the heterogeneities of the landscape. The geographical details described in the previous sections can be effectively implemented on a mesh whose generation is outlined by the following procedure:

1. Identify the region of interest in the computational domain and its main geographical features (e.g. rivers, lakes, urban areas)
2. Generate an initial grid \mathcal{T}_0 , uniformly refined over the region of interest.
3. Include elevation information from NASA SRTM data by interpolation on the grid \mathcal{T}_0 .
4. Refine \mathcal{T}_0 in the surroundings of the main geographical features identified in step (1).
5. Implement a gradient-driven refinement of \mathcal{T}_0 where the elevation gradients are steeper (like along the sides of a valley).

Basque Country Test Case

To illustrate the characteristics of our method, we simulate the dispersal of an invasive species in the territory of the Basque Country in Northern Spain. The region itself is an excellent testbed, as the heterogeneity of its landscape encompasses various types of environment from the sea shores of the Bay of Biscay along the Atlantic Ocean to the mountainous region in the interior, to the higher peaks of the nearby Pyrenees range. The computational domain Ω is the region comprised between 4°W and 1°W in Longitude, and between 42°N and the Bay of Biscay (or $44^\circ 2'\text{N}$) in Latitude (see Fig. 1, panel A). The domain Ω contains the region of interest and is triangulated according to Steps (1–5) described in the previous Section. The mesh consists of 78,832 points and 155,478 triangular elements. Note that throughout the rest of the paper co-ordinates will be expressed in Longitude and Latitude.

In the simulated scenario the invasive species is initially absent from the region of interest and is concentrated in a circle of radius 0.5 centred in the Aquitaine region in south-western France ($-1.2349, 43.8631$). From the perspective of environmental conservation, the natural interest is in the population density and speed of propagation. Some more specific quantities of interest may be the arrival time to specific, sensitive locations (like a natural reserve). As an example we consider the three major cities in Basque Country: Vitoria, Bilbao and San Sebastian. Their locations and the initial position of the invasive species are highlighted in Fig. 1, panel A, that also shows the elevation of the region. As this is an illustrative case, for numerical simplicity the impact of rivers was assumed to be negligible compared to the extended geographical heterogeneities induced by the mountain ranges. We assume a constant diffusion coefficient $\nu = 5 \times 10^{-4} \text{ deg}^2 \text{ years}^{-1}$ while the carrying capacity for the species is assumed to depend solely on elevation:

$$\gamma(\vec{x}, t) = \left(10 - \frac{z(\vec{x})}{100} \right). \quad (6)$$

The carrying capacity is thus maximum at sea level and is linearly decreasing with elevation. The species is assumed to be unable to survive at elevations greater than 1000 meters above sea level. Results of the dispersal simulation are shown in Fig. 1, panels B to F. As expected, higher densities are observed along the coastline. At the same time, the impact of elevation is evident. In particular, the barrier effect played by the Pyrénées ridge running along the 43^rd parallel significantly affects the dispersal and forces the species to move west (panel B and C) until the presence of a valley in the N-S direction allows the species to move south as well (panel D and E). Panel G in Fig. 1 shows the arrival time (t_{arrival}) of the species in the whole computational domain: t_{arrival} is defined in any point \vec{x} of the domain Ω as the time at which the population density reaches 1 i.e. $u(\vec{x}, t_{\text{arrival}}) = 1$. Black areas are inhospitable for the species due to their elevation and will not be invaded. Finally, panel H in Fig. 1 shows the temporal evolution of the population density in the three cities of interest. Naturally, San Sebastian is the first city reached by the species. Nevertheless, Bilbao is reached before Vitoria, despite the second one being closer to the initial distribution of the invasive species. Finally, the population density in Vitoria is significantly lower than in Bilbao and San Sebastian due to Vitoria's higher elevation (as can be inferred from Fig. 1, panel A).

Uncertainty Quantification

Propagating model uncertainty using arbitrary polynomial chaos. As the invasive species dispersal model is intended to be an aide in decision-taking processes for policymakers, it is necessary to quantify the model uncertainties in order for its predictions to be trustworthy. It is likely that the exact values of the coefficients in Eq. (1) are not known precisely but perhaps are known in distribution form. An arbitrary Polynomial Chaos (aPC) method is used to quantify how the variation in the values of the model coefficients affect the

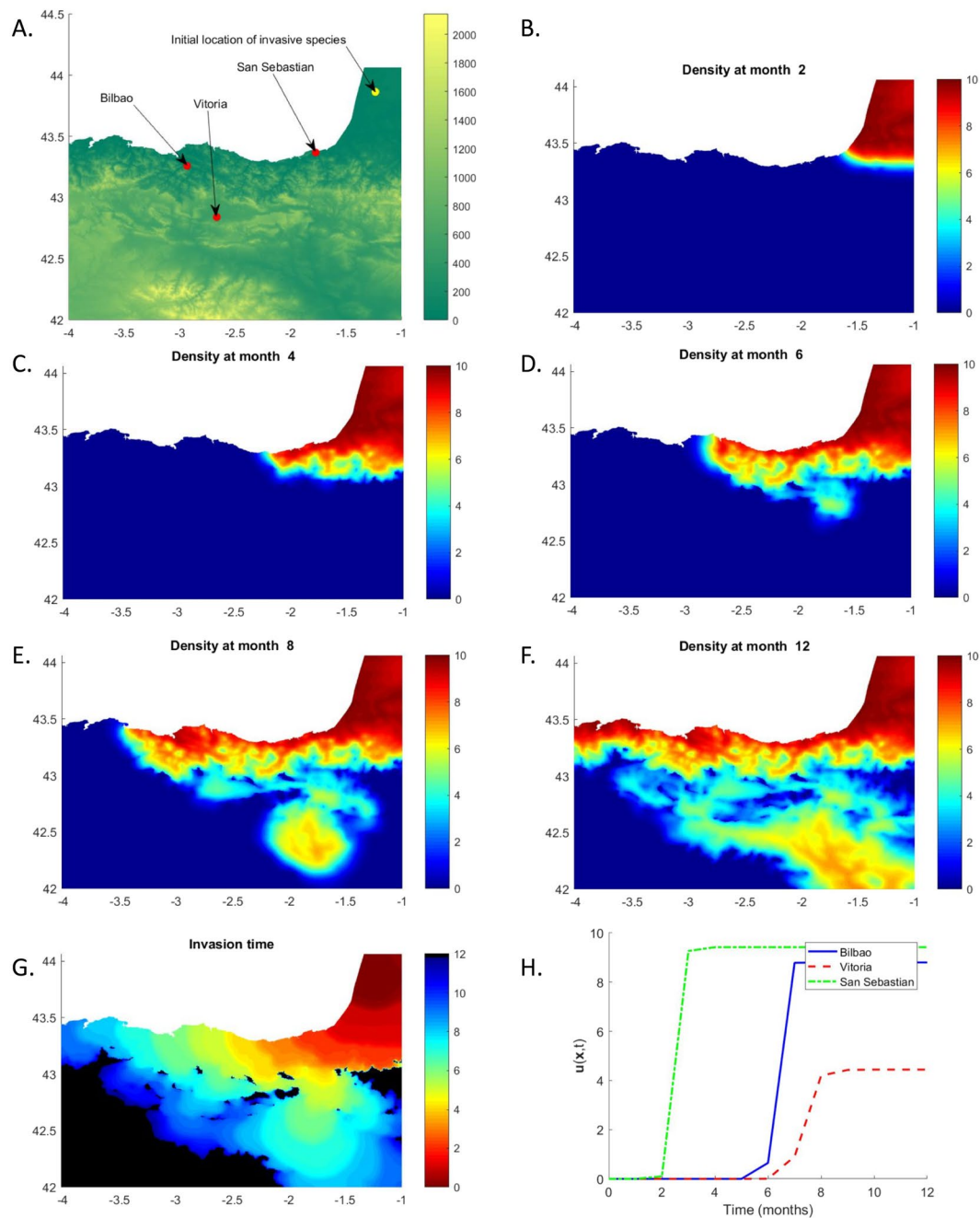


Figure 1. Basque Country test case. Panel (A) Computational domain with elevation, initial location of the invasive species and location of the three cities under study. Panel (B to F) Density of the invasive species at different times of the invasion process. The temporal dynamics highlights how extended heterogeneities due to the mountain ranges favors dispersal to the West. Panel (G) Invasion time in the whole computational domain. Black areas are inhospitable for the species and will not be invaded. Panel (H) Temporal dynamics of the population density $u(\bar{x}, t)$ for Bilbao (solid blue), Vitoria (dashed red) and San Sebastian (dot-dashed green).

predictions of the spread of the invasive species. A summary of data driven aPC is provided here, for a fuller description the reader is referred to the work of Oladyshkin⁴¹.

Data driven aPC has two advantages over the more traditional Monte Carlo sampling technique: firstly, it is far less computationally expensive than Monte Carlo sampling and secondly, the method is able to handle scarce data. The computational efficiency of the method comes from the construction of a sparse grid for sampling using Smolyak's rule. For each input distribution the optimal Gaussian quadrature points and weights are computed using the distributions' statistical moments. Repeating this process for all N_d input distributions will produce a sequence of one dimensional quadrature rules $\{U^{i_j}\}_{j=1 \dots N_d}$ with collocation points $\xi_k^{i_j}$ and weights $\omega_k^{i_j}$:

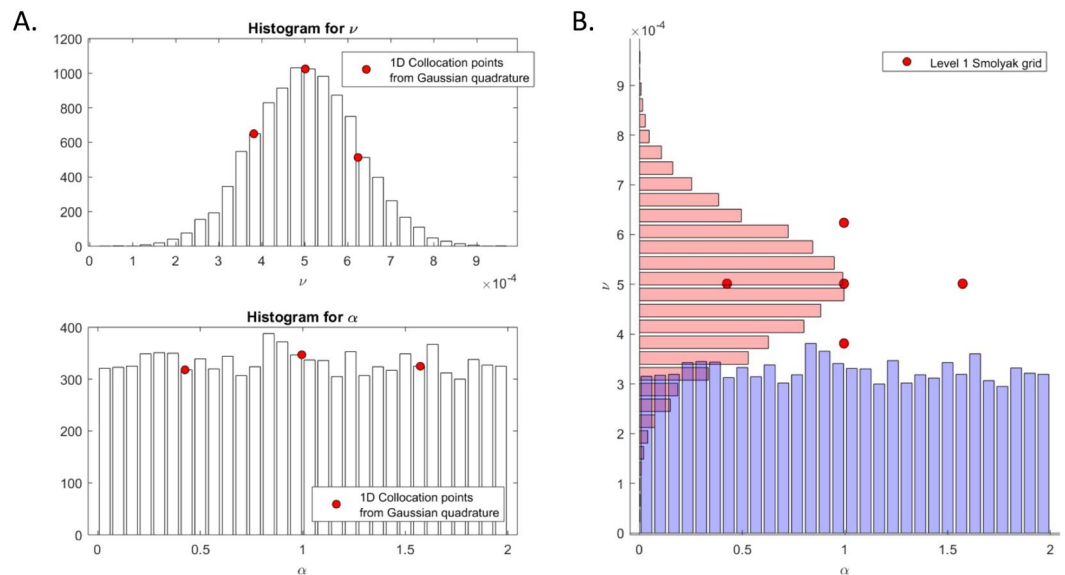


Figure 2. Panel (A) Histograms for the uncertain parameters ν and α , and the 1D collocation points for the input distributions. Panel (B) The sampling grid calculated through application of Smolyak’s rule at level 1.

$$U^{i_j} = \sum_{k=1}^{m_{i_j}} w(\xi_k^{i_j}) \omega_k^{i_j}, \tag{7}$$

where $m_{i_j} \in \{1, \dots, N_u\}$ is the maximum adaptive order for each quadrature rule and $w(\xi_k^{i_j})$ refers to a model evaluation at the collocation point $\xi_k^{i_j}$. Through the application of Smolyak’s rule it is possible to find a sparse grid to sample over:

$$A_s = \sum_{l+1 \leq |i| \leq l+N_u} (l-1)^{l+N_u-|i|} \binom{N_u-1}{l+N_u-|i|} \otimes_{k=1}^{N_u} U^i. \tag{8}$$

Increasing the level of the quadrature increases the accuracy of the result as more points are added to the grid, but at increased computational cost. Having obtained a sparse grid with N_{sp} points from Smolyak’s rule the data driven arbitrary Polynomial Chaos representation of the model output is approximated as:

$$w(\xi) \approx \sum_{k=1}^{N_{sp}} \alpha_k \psi_k(\xi), \quad \text{with} \quad \alpha_k = \frac{\sum_{i=1}^{N_{sp}} w(\eta_i) \psi_k(\eta_i) \theta_i}{\sum_{i=1}^{N_{sp}} \psi_k(\eta_i) \theta_i}, \tag{9}$$

where ψ_k refers to the k^{th} order polynomial in a family of orthogonal polynomials, while η_i and θ_i represent the sparse grid collocation points and weights. Through Monte Carlo simulation of the PCE representation of the model output it is possible to build an output probability distribution function (PDF) for the model.

Uncertainty analysis for the Basque Country test case. An uncertainty analysis was performed for the Basque Country test case. As an illustrative example we explored uncertainty in both diffusion (affecting propagation, thus the speed of the invasion) and carrying capacity (playing a key role in the settling of the species in a given location). To this end, we considered a modified version of the carrying capacity, scaled by a random parameter α :

$$\gamma(\vec{x}, t) = \alpha \left(10 - \frac{z(\vec{x})}{100} \right). \tag{10}$$

The uncertain parameters are thus the diffusion coefficient, ν , and the scaling factor α . Histograms for the two uncertain parameters were generated synthetically and are shown in panel A of Fig. 2. The histogram for ν was generated from a normal distribution $N(5 \times 10^{-4}, 1.2 \times 10^{-4})$, while α was sampled from the uniform distribution $U(0, 2)$. Having generated synthetic histograms for ν and α , data driven aPC was used to propagate the uncertainties in these parameters through the invasive species model. The aPC technique was implemented in self-developed code following the procedure described in⁴⁶. The simulation code can be made available by contacting f.montomoli@imperial.ac.uk. Panel A in Fig. 2 shows also the locations of the 1D collocation points for the two input distributions. Having found the collocation points and weights from the data, Smolyak’s rule is applied to generate a level 1 grid. In this case the resulting grid consists of 5 points, shown in panel B of Fig. 2. The invasive

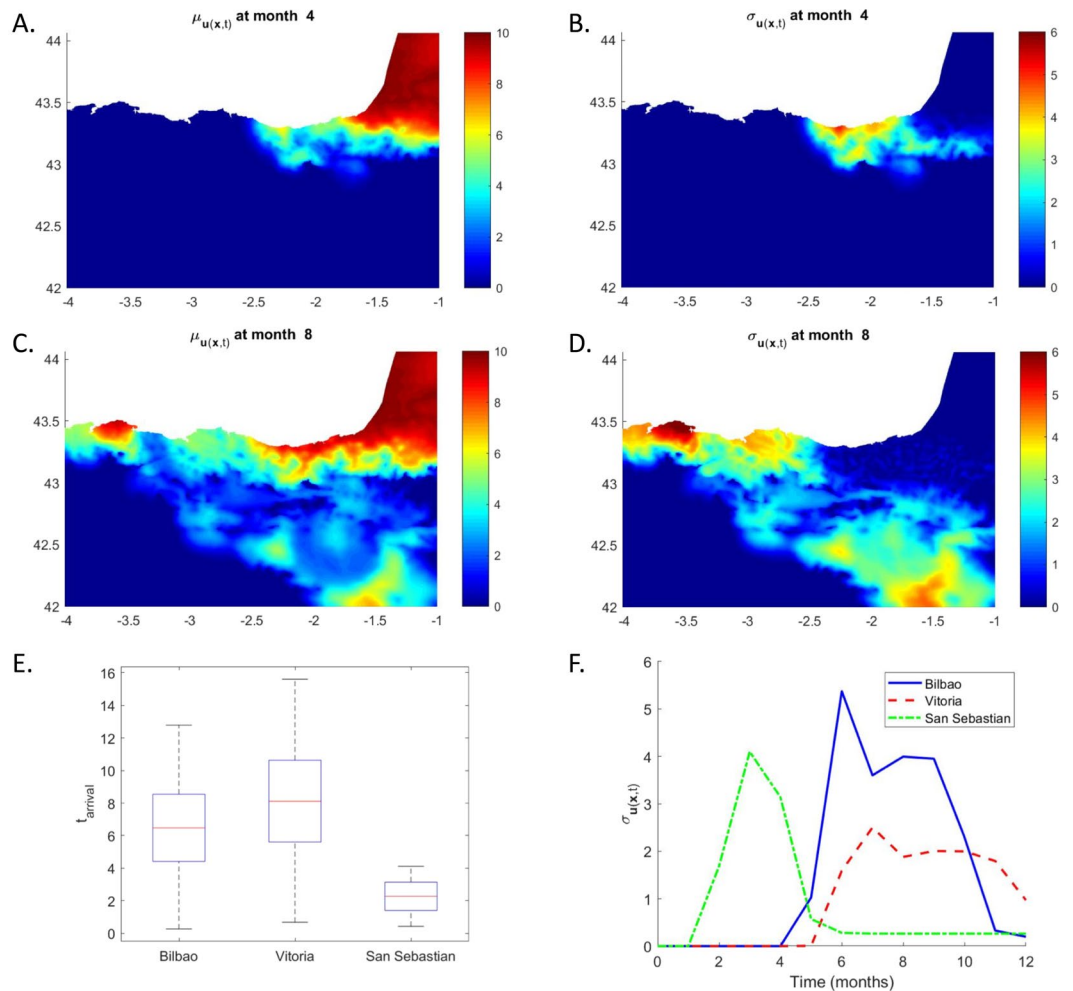


Figure 3. Uncertainty analysis for the Basque country test case. Panel (A to D) Temporal dynamics of mean and standard deviation of the population density in the whole computational domain highlights the higher level of uncertainty in the surrounding of the wavefront. Panel (E) Uncertainty in arrival time in Bilbao, Vitoria and San Sebastian. Panel (F) Temporal dynamics of the standard deviation for the population density in Bilbao (solid blue), Vitoria (dashed red) and San Sebastian (dot-dashed green) shows how uncertainty drops significantly in the wake of the propagation wavefront.

species model was evaluated at each of the points and based on the results, PDFs of $\vec{u}(\vec{x}, t)$ were calculated, at each time step, in the whole computational domain. Figure 3 collects the results. Panels A to D show the mean (μ , A and C) and the standard deviation (σ , B and D) of the PDFs on the whole region at different time steps. The bottom row of Fig. 3 show the results for Bilbao, Vitoria and San Sebastian. Panel E represent uncertainty in the arrival times of the species in the three cities, while panel F shows the temporal dynamics of the standard deviation of the population density. Panels B, D, and F highlight how uncertainty is mainly concentrated in the vicinity of the propagation front and drops significantly in its wake.

Conclusions

Predictive modeling of the spatio-temporal spread of an invasive species is a very useful tool to help environmental conservation policymakers in their decision-making process for devising appropriate countermeasures. As model parameters for invasive species are often inferred from field measurements, they are imbued with uncertainty. If the species is poorly understood then the uncertainty may be so significant that it affects the reliability of the model predictions. A continuous reaction-diffusion model for the population density of an invasive species has been presented in this paper, which has been coupled with an arbitrary Polynomial Chaos (aPC) method to assess how the uncertainty of poorly known parameters affects the predictions of the deterministic model. The coupled model has several strengths. Firstly, aPC is far less computationally expensive compared to a classical Monte Carlo sampling method. More importantly, aPC is able to handle scarce data and in this respect the proposed model is far more effective than SEPMS (which require very detailed data about the population parameters to achieve good performance)^{9–12}, and can be employed to estimate the initial stages of the invasion. Secondly, accurate geographical modeling can be easily included in the proposed framework since the continuous

reaction-diffusion model has been implemented using a Finite Element formulation. Such an approach allows the treatment of general geometries like a real geography and its actual boundaries. At the same time, it allows landscape heterogeneities to be treated in a natural way, differentiating it from other works in the literature which, relying on Finite Differences, have lacked spatial accuracy by considering species dispersal on simplified or idealised landscapes (either one-dimensional or cartesian)^{23,27,32}. The deterministic model presented here is equipped with a data-driven mesh generator that allows added refinement in the region of interest and in heterogeneities in the landscape. Finally, the individual runs of the deterministic model needed to set up the aPC procedure are not computationally demanding and can be computed on a laptop in few minutes. As a proof of concept, in this paper we demonstrate the practical applicability of the proposed method by quantifying the uncertainty in the spread of a generic invasive population in the Basque Country area in northern Spain.

For the sake of simplicity in presentation the method has been described here in basic form, but it can be easily extended in a number of ways. For instance, in our proof of concept we showed how elevation can be easily included and the carrying capacity was dependent solely on it. While the elevation is an important^{18,22,36–39} and, as has been discussed, often overlooked factor in species dispersal it is of course not the only factor. In practice the carrying capacity may be a function of a number of environmental variables. The spatial dependence of the model coefficients allows other characteristics concerning hospitality to be easily incorporated. As an example, the landscape can be differentiated between meadows, forests, urban areas and mountains, whose geographical distribution can be obtained from GIS data¹¹. Moreover, the temporal dependence of the coefficients allows for temporal variations to be accounted for in the dynamics. For instance seasonality could be modeled by periodizing relevant coefficients in the reaction-diffusion equation. Another important aspect to underline is that, for illustrative purposes, this paper considered the very simple situation of a single species invading an empty environment. However, any ecological model of competition based on systems of ordinary differential equations (like the one in Lohr *et al.*¹⁹, for instance) could be easily included in our framework. The resulting model would be a system of partial differential equations of reaction-diffusion type like the one studied in one spatial dimension by Maciel and Lutscher²³.

The presented model has some limitations as well. Firstly, although aPC is very efficient in handling scarce data, a preliminary parameter estimation for the model coefficients is needed. As population data are normally available at geopolitical or territorial levels, particular care would be required to translate them at a spatially continuous level. Secondly, although the proposed coupled model provides a probability distribution for relevant quantities of interest in species invasion, it lacks the ability of reproducing peculiar dynamics such as the stochastic fade-out characteristic of individual based models in epidemiology¹³.

Received: 16 June 2019; Accepted: 22 October 2019;

Published online: 07 November 2019

References

- Gurevitch, J. & Padilla, D. K. Are invasive species a major cause of extinctions? *Trends in Ecology and Evolution* **19**, 470–474 (2004).
- McGeoch, M. *et al.* Global indicators of biological invasion: species numbers, biodiversity impact and policy responses. *Diversity and Distributions* **16**, 95–108 (2010).
- Administration of the Basque Country Autonomous Community Department of Environment and Territorial Policy. Biodiversity Strategy of the Basque Autonomous Community 2030, http://www.euskadi.eus/contenidos/documentacion/biodiversidad2030/en_def/adjuntos/BiodiversityStrategy2030.pdf (2016).
- Tobin, P. C., Berec, L. & Liebhold, A. M. Exploiting Allee effects for managing biological invasions. *Ecology letters* **14**, 615–624 (2011).
- Hopkins, G. A., Forrest, B. M., Jiang, W. & Gardner, J. P. A. Successful eradication of a non-indigenous marine bivalve from a subtidal soft-sediment environment. *J. Appl. Ecol.* **48**, 424–431 (2011).
- Shanmuganathan, T. *et al.* Biological control of the cane toad in Australia: a review. *Animal Conservation* **13**, 16–23 (2010).
- Kaiser, B. A. & Burnett, K. M. Spatial economic analysis of early detection and rapid response strategies for an invasive species. *Resource and Energy Economics* **32**, 566–585 (2010).
- Epanchin-Niell, R. S. & Hastings, A. Controlling established invaders: integrating economics and spread dynamics to determine optimal management. *Ecology Letters* **13**, 528–541 (2010).
- Rushton, S. P., Lurz, P. W. W., Fuller, R. & Garson, P. J. Modelling the Distribution of the Red and Grey Squirrel at the Landscape Scale: A Combined GIS and Population Dynamics Approach. *Journal of Applied Ecology* **34**, 1137–1154 (1997).
- Lurz, P. W. W. *et al.* Predicting grey squirrel expansion in North Italy: a spatially explicit modelling approach. *Landscape Ecology* **16**, 407–420 (2001).
- Tattoni, C. *et al.* Modelling the expansion of a grey squirrel population: implications for squirrel control. *Biological Invasions* **8**, 1605–1619 (2006).
- Jones, H., White, A., Lurz, P. & Shuttleworth, C. Mathematical models for invasive species management: Grey squirrel control on Anglesey. *Ecological Modelling* **359**, 276–284 (2017).
- Macpherson, M. F. *et al.* Incorporating habitat distribution in wildlife disease models: conservation implications for the threat of squirrelpox on the Isle of Arran. *Anim. Conserv.* **19**, 3–14 (2016).
- Catford, J., Vesik, P., White, M. & Wintle, B. Hotspots of plant invasion predicted by propagule pressure and ecosystem characteristics. *Diversity and Distributions* **17**, 1099–1110 (2011).
- Gassó, N., Pino, J., Font, X. & Vilà, M. Regional context affects native and alien plant species richness across habitat types. *Applied Vegetation Science* **15**, 4–13 (2012).
- Campos, J. A. *et al.* Climate and Human Pressure Constraints Co-Explain Regional Plant Invasion at Different Spatial Scales. *PLoS One* **11**, 1–19 (2016).
- Townsend Peterson, A. Predicting the Geography of Species' Invasions via Ecological Niche Modeling. *The Quarterly Review of Biology* **78**, 419–433 (2003).
- Haran, J., Roques, A., Bernard, A., Robinet, C. & Roux, G. Altitudinal Barrier to the Spread of an Invasive Species: Could the Pyrenean Chain Slow the Natural Spread of the Pinewood Nematode? *PLoS One* **10**, 1–25 (2015).
- Lohr, C. *et al.* Modeling dynamics of native and invasive species to guide prioritization of management actions. *Ecosphere* **8**, 1–19 (2017).
- Meier, E. *et al.* Space matters when defining effective management for invasive plants. *Diversity and Distributions* **20**, 1029–1043 (2014).
- Guisan, A. & Zimmermann, N. E. Predictive habitat distribution models in ecology. *Ecological Modelling* **135**, 147–186 (2000).
- Dullinger, S. *et al.* Extinction debt of high-mountain plants under twenty-first-century climate change. *Nature Climate Change* **2**, 619–622 (2012).

23. Maciel, G. A. & Lutscher, F. Movement behaviour determines competitive outcome and spread rates in strongly heterogeneous landscapes. *Theoretical Ecology* **11**, 351–365 (2018).
24. Smith, D. L., Lucey, B., Waller, L. A., Childs, J. E. & Real, L. A. Predicting the spatial dynamics of rabies epidemics on heterogeneous landscapes. *Proceedings of the National Academy of Sciences* **99**, 3668–72 (2002).
25. Fisher, R. A. The wave of advance of advantageous gene. *Ann Eugenics* **7**, 355–369 (1937).
26. Acevedo, M., Marcano, M. & Fletcher, R. Jr. A diffusive logistic growth model to describe forest recovery. *Ecological Modelling* **244**, 13–19 (2012).
27. Bonneau, M., Johnson, F. & Romagosa, C. Spatially explicit control of invasive species using a reaction diffusion model. *Ecological Modelling* **337**, 15–24 (2016).
28. Fritts, T. & Rodda, G. The role of introduced species in the degradation of island ecosystems: A Case History of Guam. *Annu. Rev. Ecol. Syst.* **29**, 113–140 (1998).
29. Sax, D., Gaines, S. & Brown, J. Species Invasions Exceed Extinctions on Islands Worldwide: A Comparative Study of Plants and Birds. *The American Naturalist* **160**, 766–783 (2002).
30. Steadman, D.W. *Extinction and Biogeography of Tropical Pacific Birds* (Univ. of Chicago Press, 2006).
31. Formaggia, L., Saleri F. & Veneziani, A. *Solving Numerical PDE's* (Springer, 2012).
32. Miller Neilan, R. & Lenhart, S. Optimal vaccine distribution in a spatiotemporal epidemic model with an application to rabies and raccoons. *J. of Math Anal App* **378**, 603–619 (2011).
33. Hooten, M. B. & Wikle, C. K. A hierarchical Bayesian non-linear spatio-temporal model for the spread of invasive species with application to the Eurasian Collared-Dove. *Environ Ecol Stat* **15**, 59–70 (2008).
34. Arab, A. Hierarchical spatio-temporal models for environmental processes. PhD Thesis, University of Missouri–Columbia (2007).
35. Keller, J., Gerardo-Giorda, L. & Veneziani, A. Numerical simulation of a susceptible-exposed-infectious space-continuous model for the spread of rabies in raccoons across a realistic landscape. *J. Biological Dynamics* **7**(Suppl 1), 31–46 (2013).
36. Averett, J. P. *et al.* Non-Native Plant Invasion along Elevation and Canopy Closure Gradients in a Middle Rocky Mountain Ecosystem. *PloS one* **11**, 1–24 (2016).
37. Steinbauer, M. J. *et al.* Plant invasion and speciation along elevational gradients on the oceanic island La Palma, Canary Islands. *Ecol. Evol.* **7**, 771–779 (2017).
38. Pauchard, A. *et al.* Non-native and native organisms moving into high elevation and high latitude ecosystems in an era of climate change: new challenges for ecology and conservation. *Biological Invasions* **18**, 345–353 (2016).
39. Leroux, S. *et al.* Mechanistic models for the spatial spread of species under climate change. *Ecological Applications* **23**, 815–828 (2013).
40. Ern, A. & Guermond, J. L. *Theory and Practice of Finite Elements* (Springer, 2006).
41. Oladyshkin, S., Class, H., Helmig, R. & Nowak, W. A concept for data-driven uncertainty quantification and its application to carbon dioxide storage in geological formations. *Adv. Water Resour.* **34**, 1508–1518 (2011).
42. Okubo, A. & Levin, S. A. *Diffusion and ecological problems: modern perspectives* (Springer, 2013).
43. Salsa, S. *Partial differential equations in action. From modelling to theory* (Springer, 2008).
44. Shan, H., Li, J. & Chaoqun, L. Direct numerical simulation of flow separation around a NACA 0012 airfoil. *Computers & fluids* **34**, 1096–1114 (2005).
45. Saad, Y. *Iterative Methods for Sparse Linear Systems* (PWS, 1996).
46. Pepper, N., Montomoli, F. & Sharma, S. Multiscale uncertainty quantification with arbitrary polynomial chaos. *Computer Methods in Applied Mechanics and Engineering* **357**, 1–20 (2019).

Acknowledgements

N.P. is funded jointly by the EPSRC and Airbus through iCASE, voucher number 17000099. L.G.G. is supported by the Basque Government through the BERC 2018–2021 program, and the Spanish State Research Agency (AEI) through the BCAM Severo Ochoa excellence accreditation SEV-2017-0718. L.G.G. is also funded jointly by AEI and Feder through the grant RTI2018-093416-B-I00 MULTIQUANT. The authors wish to thank the anonymous reviewers for their valuable comments to improve the manuscript.

Author contributions

N.P. wrote the main manuscript text and carried out the uncertainty analysis. L.G.G. and F.M. jointly developed the distributed model for species dynamics and its numerical implementation, including the data-driven mesh generator. All authors reviewed the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information is available for this paper at <https://doi.org/10.1038/s41598-019-52763-9>.

Correspondence and requests for materials should be addressed to N.P.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2019