Incorporating the geometry of dispersal and migration to understand spatial patterns of species distributions

Gimenez, Luis

Ecography

DOI:
10.1111/ecog.03493

Published: 01/04/2019

Peer reviewed version

Incoporating the geometry of dispersal and migration to understand spatial patterns of species distributions

Luis Gimenez

School of Ocean Sciences, Bangor University, LL50 5AB, Menai Bridge, Anglesey, UK

Corresponding author: Luis Gimenez, School of Ocean Sciences, Bangor University, LL50 5AB, Menai Bridge, Anglesey UK. E-mail: l.gimenez@bangor.ac.uk

Decision date: 10-Sep-2018

Abstract

Dispersal and migration can be important drivers of species distributions. Because the paths followed by individuals of many species are curvilinear, spatial statistical models based on rectilinear coordinates systems would fail to predict population connectivity or the ecological consequences of migration or species invasions. I propose that we view migration/dispersal as if organisms were moving along curvilinear geometrical objects called smooth manifolds. In that view, the curvilinear pathways become the "shortest realised paths" arising from the necessity to minimise mortality risks and energy costs. One can then define curvilinear coordinate systems on such manifolds. I describe a procedure to incorporate manifolds and define appropriate coordinate systems, with focus on trajectories (1D manifolds), as parts of mechanistic ecological models. I show how a statistical method, known as "manifold learning", enables one to define the manifold and the appropriate coordinate systems needed to calculate population connectivity or study the effects of migrations (e.g. in aquatic invertebrates, fish, insects and birds). This approach may help in the design of networks of protected areas, in studying the consequences of invasion, range expansions, or transfer of parasites/diseases. Overall, a geometrical view to animal movement gives a novel perspective to the understanding of the ecological role of dispersal and migration.

Introduction

Space plays a central role in ecological processes, driving patterns of population abundance, community structure and ecosystem function. There is now a significant body of work showing that ecological processes operate on populations and communities at various spatial scales, leading to scale-dependent patterns of structure (e.g. patchy distribution of organisms) from the scale of metres to those encompassing meta-populations and communities (Wiens 1989, Levin 1992, Wiens and Bachelet 2010). In populations occupying fragmented landscapes (e.g. animals living in forest patches separated by cultivated grassland; marine benthic invertebrates living in rock shores separated by sandy beaches), large scale ecological patterns can be driven by migration/dispersal among patches in addition to mortality and reproduction. There is also a significant amount of work highlighting the importance of distance-dependency in the intensity of interactions. Movement, migration and dispersal across space are usually distance-dependent, meaning that nearby patches exchange individuals more easily (i.e. they are better connected) than those located far away. In addition, neighbours usually interact with target individuals more frequently than those located outside the neighbourhood. Distance-dependency in the intensity of interactions can promote coexistence between predators and prey (Tilman and Kareiva 1997, Dieckmann et al. 2000), and the formation of regular patterns of distribution in many types of ecosystems (Rietkert and Van de Koppel 2008). In addition, patterns arise as the consequence of dispersal limitations imposed by habitat fragmentation and landscape features (Hanksi 1999, Holyoak et al. 2005, Massol et al. 2011). Landscape and seascape structures promote spatial patterns in population and communities because they provide environmental heterogeneity and also because they determine patterns of migration or dispersal across the landscape (Stamps et al. 1987, McIntyre and Wiens 1999, Massol et al. 2011). Organisms may connect patches by dispersing in all directions across the landscape (e.g. across cultivated land) or may follow specific migratory routes. An important point is that the pathways of migration/dispersal do not always take the shape of straight lines across the habitat matrix (i.e. understood here as the habitat that animals cross) or follow trajectories solely defined by the earth curvature (Greenberg et al. 2011). In some cases, such movements occur through a network of "minimum cost corridors" or "least cost paths" (e.g. Adriaensen et al. 2003, Estrada-Peña 2003, Greenberg et al. 2011). In others, movement has been abstracted as a walk on a fractal space (Crist et al. 1992, Scheiner

1994 chap 5, Edwards, et al. 2007). In addition, there is now a growing body of theory suggesting that, at large scales, marine organisms disperse or are transported by currents along curves and meanders as a consequence of the curl of eddies or curvilinear coastlines (White et al. 2010, Butler et al. 2011, Robins et al. 2013). Curvilinear trajectories also occur in insects migrating on wind fields (Chapman et al. 2010), by migrating birds (Mandel et al. 2011, Trerweiler et al, 2014) and most likely by organisms dispersing along rivers (Boedeltje et al. 2003, Nathan 2008). The curvilinear nature of the migration pathways are usually the consequence of behavioural responses of migrants to e.g. landscape topography, predators or food.

The problem imposed by curvilinear trajectories is that they will, in principle call into question the use of Euclidean distances (or distances based on geographic coordinates) as the metric to quantify connectivity or distance-dependent interactions (Greenberg et al. 2011). For instance, for those cases, connectivity between local populations is not well described by straight lines. Indeed, in the marine environment, straight lines (or arcs) defined as Euclidean distances (or geographical distances based on latitude and longitude) do not explain patterns of connectivity (Moritz et al. 2013, D'Aloia et al. 2015) or genetic population structure (Alberto et al. 2011) because the actual "oceanographic distances" between populations respond to highly curved trajectories (White et al. 2010). This is an important issue in the context of conservation, at the time of defining spatial distances between protected areas (Botsford et al. 2001, D'Aloia et al. 2015).

The idea proposed here is to use concepts of differential geometry of smooth surfaces in order to obtain appropriate metrics of (curvilinear) distances that are ecologically meaningful for understanding patterns of species distribution. I focus on curvilinear trajectories of organisms, but the same concepts can be applied to transport of materials and its role in energy flow and ecosystem processes. In the first section, I organise these ideas in a framework, based on concepts of differential geometry that enable the definition of curvilinear distances. Then, I use a simple example to introduce these concepts as a way to show how alternative metrics of distance (curvilinear or Euclidean) can be defined and tested. In the third section, I generalise the procedure of testing whether different metrics of distance are appropriate descriptors of distance-dependent ecological processes. This procedure involves the tests of alternative definitions of distance, through standard spatial statistical methods (Isaaks and Srivastava 1989, Plant 2012); by extension, alternative hypothetical dispersal/migration mechanisms are tested.

In the fourth second section, I discuss statistical methods available to estimate curvilinear distances in real life situations. To the best of my knowledge, an approach such as the one proposed here has not been done before (Supplementary material appendix: Section 1). The exception are cases where distances are calculated on the earth surface, considered as a 2D curved space (using e.g. a spherical coordinate system), but here I consider highly curved trajectories.

2. Conceptual framework

The fundamental concepts to be discussed here are "distance", "length" and "space"; the discussion is based on mathematical elaborations within the field of differential geometry (Grinfeld 2010, Peters 2010, Pressley 2012). Importantly, the calculation of *distances* among objects (e.g. habitat patches) in a given *space* (e.g. the flat, Euclidean space) is usually done with reference to a *coordinate system* (e.g. a Cartesian). In the Euclidean space (and in any other curvilinear space), we can define any coordinate system to describe the position of any two habitat patches (details in Supplementary material appendix Section 2: Fig. A1). The coordinates of the patches will vary with the coordinate system, but the distance between them will remain constant irrespective of the coordinate system considered; distances are said to be invariant to a coordinate system transformation. However, distances differ if we view the habitat patches as placed in different types of spaces (Supplementary material appendix Section 2: Fig. A1). Hence, the definition of distances is tied to the characteristics of the space where the habitat patches are placed.

The use of Euclidean distances in ecology implies the assumption of animals moving freely in the flat physical space; by using geographic coordinates (latitude, longitude) the implication is that organism move in a space defined by the curvature of the earth surface. However, here I am referring to cases where animal movement follows curvilinear trajectories defined by their interaction with the landscape. In such situations, migration may be viewed as if organisms were moving on or along curvilinear geometrical objects, defined in geometry as *smooth manifolds* (Fig.1). Smooth manifolds are smooth surfaces that locally resemble the Euclidean (=flat) space but are curved at large scales. Hence, at a small scales, one can map points on a manifold using a coordinate system defined of the Euclidean space. In addition, for these manifolds, one can produce one-to-one mapping of each point of the manifold into a chart. The sphere and different types paraboloids (Fig 2) are examples of 2-dimentional (2D)-smooth manifolds. The earth can be represented as near spherical

manifold; for this case, the Mercator projection maps each point of the manifold (the earth surface) into a plane. Smooth curves without self-crossings are examples of 1D manifolds of the type discussed here. Smooth manifolds generalise the concept of spaces, from Euclidean to curved spaces; i.e. a plane or the 3D Euclidean space are also manifolds.

The fact that manifolds can be considered locally flat but regionally curved is consistent the concept of scale-dependence in ecology applied to dispersal and migration. At short distances, the paths of dispersal may be well *approximated* straight lines (for the cases discussed here), which implies the assumption that organisms move in a Euclidean space. However, at large scales, interactions between landscape features and organism lead to curvatures that play a significant role in defining distances among habitat patches. For example, at the scale of a few hundreds of meters, the path of a flying bird or a marine organism drifting along a current may look straight. However at scales of hundreds of kilometres such paths are likely to be curved: birds usually avoid (and fly around) habitats where risks of mortality are high; marine animals drift around eddies or other types of loops produced by the rotation of water masses.

Another important concept is related to the interpretation of migration/dispersal pathways. The paths followed by organisms may be viewed as the "shortest realised paths" in spite of the fact that they are actually curved. This interpretation is valid (ecologically) because such paths, shaped by behaviour and the landscape, are likely to minimize energy costs or risks of mortality. For example, an organism drifting in the sea follows curvilinear pathways because alternative routes (swimming against currents) are energetically demanding or physical impossible. Birds, may avoid crossing large water masses or flying above habitats with little food availability due to high risk of mortality. Importantly, on smooth manifolds, the shortest distance between two points are segments of special curves called the geodesics. The concept of geodesics generalises the concept of straight lines for curvilinear spaces (Fig. 2) i.e. in a Euclidean space, the geodesics are straight lines; in a curvilinear manifold the geodesics are curves.

Taking into account the concept of manifold and geodesics we can view migration as if individuals were moving on the surface of a curvilinear manifold, arising as the consequence of the interactions of individuals with the landscape. In addition, we can view the pathways of migration as the "shortest realised paths" within the manifold and therefore as sections of the manifold geodesics. For example, if we abstract the earth as

a sphere, in the absence of wind patterns, rotation effects or unsuitable habitats, the shortest path followed by a migrating bird would be the one defined by a great circle, i.e. a circle centred at the centre of the earth. This is because the great circles are the geodesics of the sphere. However, once we accept that animal migration occurs along manifolds arising from the presence of e.g. wind patterns and unsuitable habitat, then the resulting curved trajectories become the geodesics of that manifold.

On the manifolds, distances between habitat patches (or organisms) may be calculated on the geodesics, with reference to a curvilinear coordinate system (Fig. 1). Here, I call such distances as "ecologically based geodesic distances" (EGDs), because they constitute the geodesic of a manifold arising from ecological processes (interactions between organisms and the surrounding habitat). Given this definition, the Euclidean or geographic distances between two patches (or organisms) become particular cases of EGDs whereby animal movement between patches is not constrained by e.g. current patterns or unsuitable habitat. Because the Euclidean and geographic distances are cases of EGDs, this framework enables the construction of alternative hypotheses (each represented by a different EGD) about how ecological processes occur in space. Thus, each metric of distance becomes a testable hypothesis. The issue of hypothesis testing is further elaborated in the following two sections.

3. Applying geometrical concepts to a simple example

Consider the following thought experiment: imagine a regional coastline (Fig. 3a) harbouring evenly spaced populations of organisms connected by individuals migrating along the habitat matrix. The organism in question, terrestrial or marine (e.g. the latter dispersing during the larval phase), is invading the bay from the NE, spreading to the S and W. As time passes, colonising stages invade habitat patches so that populations near the source are larger than those located far away. In the bay, two hypothetical mechanisms of dispersal are possible, either dispersal through natural means or human-mediated dispersal. The dispersal through natural mechanism is constrained to occur along the coast: for the case of the terrestrial organism, this would be realistic if individuals are not capable of swimming or flying, at least for such long distances characterising the bay. For marine organisms, hydrodynamic conditions may constraint larval transport to occur only along the coastline; for instance, coastal fronts located at shelf breaks may restrict cross-shelf dispersal or larval behaviour may promote retention near the coast (e.g. Marta-Almeida 2006), but enable along shore transport. The human-

mediated dispersal occurs through maritime transport: for instance, terrestrial animals may hitchhike on ships; marine invertebrate larvae may be trapped in ballast water contained in ships hulls and discharged upon arrival to a new coastal site (Seebens et al. 2013) or organisms may colonise boat hulls (Carlton and Hodder 1995). Assume that maritime transport occurs in straight lines as ships move all across the bay, but transport is more frequent along shorter than longer distances. Hence, in this scenario, the alternative hypothetical dispersal mechanisms differ in the characteristics of the manifolds. For the natural dispersal mechanism, the manifold is a one-dimensional (1D) geometrical object whose shape is given by the coastline; from the standpoint of dispersing organism, the shortest realised path between two inhabitable habitat patches, the geodesics, consist of a curve that will determine the EGD between habitat patches. For the maritime transport mechanism, the manifold is a 2D flat space (for simplicity, earth curvature is ignored) and the geodesics are straight lines, and the EGDs are made of Euclidean distances among habitat patches.

Importantly, the alternative EGDs, derived from migration along each manifold, predict a different pattern of distribution. Hence, the incorporation EGDs into statistical procedures would enable determination of which mechanism is more consistent with spatial patterns. If the coastline were a straight line, one would expect a monotonic decrease in abundance along the geographical coordinates, as individuals disperse; such a pattern would be evaluated, for instance, using a simple regression model estimating abundance as a function of the linear distance defined by the geographical coordinates; i.e. that axis constitutes a convenient coordinate system. Following this argument, I choose a convenient coordinate system defined on the manifolds in order to quantify the spatial changes in the pattern of abundance of the target species, i.e. a Cartesian system for the maritime transport mechanism and a curvilinear system for the natural dispersal mechanism. The choice of the coordinate systems enables the use of standard statistical methods of hypothesis testing or model selection to determine which mechanism is more consistent with the patterns.

I now illustrate the steps described above, simulating a process of model selection in two different sets of scenarios where either the natural dispersal or the maritime mechanisms are the true mechanisms explaining the patterns of distribution of the invading space. In the scenarios where the true mechanism is the natural dispersal (EGD is curvlinear), the following key question are addressed: given a curvilinear pattern of dispersal, what is the performance of a model fitted using linear combinations

of the Cartesian coordinates? Is there any advantage in using this "manifold approach" as compared with sticking to a flat space and a Cartesian (or other rectilinear) coordinate system? Notice that this is not a circular argument: we know that the true model is the one based on curvilinear migration but we want to determine how well a model based on Euclidean distances will perform. This will tell us if an observer, not knowing the true model, will be able to use the "manifold approach" to infer the mechanism of migration.

I consider three different coastlines (Fig. 3b) as a way to evaluate the importance of the degree of landscape curvature in how good predictions are made based on geographical coordinates as compared to the curvilinear coordinate system. As in a real life scenario, the data consist of the abundance of target organism at each habitat patch, the position of each patch with respect to the Cartesian coordinates and the position in relation to the curvilinear coordinate system. The details of methods used to produce the data are given in the Supplementary material appendix: Section 3.1: Briefly, the coastline is defined as a parabola, embedded in a Euclidean space with a Cartesian coordinate system, $C(x,y)$. The equation of the parabola has a parameter $\sigma > 0$ that controls the coastline curvature, such that open coasts have large σ-*values* (Fig.3b). Along the coastline, habitat patches were defined and located according to the Cartesian coordinates. The curvilinear coordinate system, $S(z)$, was defined on the parabola as the distance along the coast. The abundance of the dispersing species, N_i at a patch i was modelled as an exponential decaying function of the spatial coordinates of each habitat patch plus normally distributed errors; the exponential decay is governed by the parameter β. For each combination of values of σ (=0.5, 1 and 2) and β (=0.1, 0.2, 0.3), I simulated 1000 patterns of abundance distributions (Fig 3c,d, summarised in fig. 4 and Supplementary material Appendix Section 3.2 figs A2-A3) and carried out the corresponding 1000 statistical tests (see below). The differences in the patterns of distribution (averaging the 1000 simulations) arising from the different mechanisms are clearly seen by comparing figures 3c and 3d. While the natural transport mechanism (Fig 3c) predicts the lowest abundance at the NW extreme of the bay, the maritime transport (Fig 3d) results in a rise in abundance at that extreme. The rise of abundance occurs because the linear distances between the NW and NE extremes, are smaller than linear distances between NE and S of the bay (see also Fig 3a).

Model fitting was made with general least square models and comparisons were made with ΔAIC (Supplementary material appendix: Section 3.1), focusing on the comparison

between the true model and three alternative models based on Cartesian coordinates (Table 1). The hypothetical mechanisms (Table 1), associated manifolds and coordinate systems are defined as follows: (1) alongshore dispersal, the predictor is the arc length of the parabola describing the shoreline. (2) Dispersal occurring along the x-dimension (the only predictor is x). (3) Dispersal across the bay, over a 2D flat space (4) Dispersal depending on the Euclidean distance; this model implies a 2D flat space where the predictor is the Euclidean distance.

The overall results (Table 2) show that models based on Cartesian coordinate systems perform as good as the true model only in the case of very open coasts ($\sigma = 2$): in that case ΔAIC are approaching zero. However, even on slightly curved coastlines ($\sigma = 1$), these models do not fare as well as the one based on the curvilinear coordinate system. Plots of the average, standard error and confidence intervals of the distribution based on all the 1000 calculated ΔAIC (e.g. Fig. 4) also highlight the low performance of models based on Cartesian coordinate systems unless the coastline is wide open. Hence, even for slightly curved migration pathways it is unlikely that an observer would fit a model based on linear combination of Cartesian coordinates with as good performance as that fitted on the curvilinear coordinates.

The poor performance of models based on Cartesian coordinates occurs because such models do not predict a decrease in abundance along the bay (as in Fig 3c). For instance, the model based on Euclidean distances would predict a rise in abundance towards the NW extreme (as shown in Fig. 3d) and the one based on x-coordinates underestimates the decrease towards the NE (Supplementary material Appendix: Section 3: Table A1). It is likely that fitting a complex quadratic or cubic polynomial or a spline function would give an acceptable model fit. However, it would be difficult to interpret the parameters of such complex (and less parsimonious) models, in ecological terms. By contrast, by defining alternative manifolds, as representations of the alternative mechanisms (e.g. maritime transport vs natural dispersal), one gives a mechanistic explanation to the patterns and hence gives ecological meaning to the decay coefficient. The mechanistic explanation exists because the shape of the manifold is associated with, for example, specific behaviours and landscape features. Hence, the decay coefficient quantifies the decrease is abundance per unit of length with respect to a specific manifold (representing a specific dispersal mechanism).

The set of scenarios where connectivity is driven by the maritime transport mechanisms enables the test of curvilinear distance. This is based on the same type of simulation

explained above (see Supplementary material Appendix: Section 3.2 Table A2**)**, but now with spatial patterns as driven by Euclidean distances. The results show that the natural mechanism (=curvilinear distance) is rejected as a mechanism explaining the spatial patterns of abundance (see Supplementary material Appendix: Section 3.2**,** fig. A4 for spatial patterns of abundance). Again, and in consistence with logic, only when the bay is characterised by a low curvature would both models show similar explanatory power. Therefore, curvilinear distances are a testable model as much as Euclidean distances.

4. Hypothesis testing/model selection

Based on the idealised case of the previous section, one can derive a general procedure that enables to test whether flat or curvilinear spaces, and hence Euclidean or arc-distances, correctly represent EGDs among organisms or habitat patches. In doing so, we test for alternative mechanisms explaining such patterns because the distances are defined from the EGDs, which in turn reflect the mechanisms driving dispersal.

The general procedure consists of three steps: (1) The formulation of hypotheses about mechanisms shaping the pathways of dispersal or migration. In this step, a unique manifold must be assigned to each specific mechanism: this makes the manifold geodesics (and hence the distances) unique to each specific hypothesis and therefore testable. (2) The choice of a convenient coordinate system on each manifold. This step enables the use of spatial statistical methods for hypothesis testing or model selection. For a 1D curvilinear alternative (Fig 3), the coordinate system will be placed along the pathway of dispersal/migration, but for a 2D space more than one coordinate systems may be defined. (3) The application of standard statistical methods (e.g. regression, generalised linear modelling) incorporating the spatial position of the habitats according to the chosen coordinate systems. The mode of testing will vary, according to the question. In the example of the previous section, the coordinates were incorporated in the model as predictors, but they may be instead included in the error term in order to consider potential spatial autocorrelation effects. Overall, this general procedure enables the test of different EGDs and by extension alternative hypothetical mechanisms driving patterns of distribution.

5. Constructing alternative hypotheses in with real world scenarios

The previous section summarised the sequence of steps needed to use manifolds starting with the formulation of hypotheses and ending in the application of linear statistical methods. Euclidean distances, as one of the alternative hypothetical EGD, is defined from Cartesian coordinate systems applied on a landscape. The difficult point, which is addressed in this section, is to "find" (= "approximate" in the statistical language) the curvilinear manifolds and the coordinate systems that are going to provide the alternative hypothetical definitions of EGDs.

The type of data used to approach the manifolds will obviously depend on the specific case. For example, manifolds may be approximated from the spatial distribution of habitat patches (e.g. fragmented forests in a cultivated landscape, islands in the sea), topographic structures in the landscape (e.g. rivers, mountains), hydrographic structures in the seascape (e.g. eddies or frontal zones), or position of organisms as they migrate. The geometry of the spatial patterns of habitats of organisms may lead to the approximation 1D manifold, a curve, when e.g. habitats/organisms are mapped on a plane or a 2D curved surface if they are mapped in three dimensions.

In order to understand what is required to approximate a manifold and a coordinate system, it is convenient to recall that Euclidean spaces constitute flat manifolds where orthogonal coordinate systems may be obtained through principal component analysis (PCA). In ecology, PCA is used as a method of dimension reduction (Legendre and Legendre 1998) but the components constitute a rotated and orthogonal coordinate system. This line of thinking shows that we require an extension of PCA, enabling one to approximate the shape of a curvilinear manifold and define a coordinate system. As recognised by Pearson (1901), linear regression methods are not appropriate to derive a coordinate system. Therefore, GAMs or non-linear regression techniques, expressing e.g. the *y-coordinate* as a function of the *x-coordinate*, $y=f(x)$, would not be appropriate either.

The extension of PCA to curvilinear spaces is called "manifold learning", a type of unsupervised machine learning technique (Gorban et al. 2008, Einbeck et al; 2010); it includes two types of method, the nonlinear PCA (e.g. Scholz et al. 2008) and principal manifold techniques (e.g. Gorban and Zinovyev 2008). Some have been used to represent objects in space ranging from molecules (Gorban and Zinovyev 2008) to anatomical structures (Failmezger et al. 2013). Manifold learning techniques have also been used as an ordination method to study patterns of gene expression (Gorban and Zinovyev 2010) and in climate science (Hsieh et al. 2006). In what follows, I apply one

of these techniques, non-linear PCA (NLPCA: Scholz et al. 2010) to seven ecological cases (details of methods in Supplementary material Appendix Section 4).

The first two examples (Butler et al. 2011; Ayata et al. 2010) are based on paths of dispersal of marine invertebrate larvae obtained from numerical models. Although not based on real data, these examples are useful to introduce the approach based on manifolds, and visualise the performance of NLPCA when migration pathways are highly curvilinear. Butler et al. (2011) modelled the dispersal of spiny lobster larvae (*Panulirus argus*) in Yucatan bay (Caribbean Sea) with respect to spawning sites; dispersing individuals form a dispersal plume; I use one such map to construct a manifold (Fig. 5a). I simulated the dispersal plume that would capture the convoluted shape of the dispersal pathways found by Butler et al. (2011) and the pattern of connectivity between 10 of the sites of larval release, located along the shore. The curvilinear coordinate system is defined on the manifold approximated through NLPCA. In addition, I derived a second coordinate axis, z_2, based on Euclidean distances between the projection of each point in the manifold (triangles in Fig 5a) and the position in the Euclidean space (details in Supplementary material Appendix Section 5). The graphical representation of the new coordinate system, $S(z_1,z_2)$, shows that geographical distances between sites do not always correspond to those produced by the manifold, but instead would represent an average oceanographic distance between sites. If model predictions of Butler et al. (2011) are correct, from the perspective of a dispersing larva, the shortest distance between spawning populations of spiny lobsters would be better defined as those given by the curvilinear manifold (Fig. 5b). At this stage, one would use the two coordinate system, $S(z_1,z_2)$ and $C(x,y)$, to construct separate statistical models and test which metric of distance (curvilinear or Euclidean) and, by extension, which mechanism of dispersal (free random dispersal vs larval transport) is more likely to explain metapopulation connectivity.

There may be situations enabling the test of several models of curvilinear dispersal along with the model of dispersal based on Euclidean distance. For instance, Ayata et al. (2010) found that different larval behaviours lead to dispersal plumes of different shapes. The manifolds (Figs 5c-e) correspond to the southernmost kernel modelled in the study area (Bay of Biscay). The one corresponding to passive transport (Fig. 5c) shows the influence of hydrodynamic structures on the shape of the dispersal plumes and differs from those corresponding to the models including animal migration (ontogenic and diel larval vertical migration). In this case, one would apply model

selection (or hypothesis testing) to four models, three of them based on the coordinate systems defined each manifold, plus a model based on a Cartesian coordinate system. In the following five examples (Fig. 6), I use actual data and focus on approximating manifolds needed to derive the curvilinear coordinate system (the Z_2 coordinate derived in Fig 5b is not longer considered). Trierweiller et al. (2014) studied patterns of migration of Montagu's harriers (*Circus pygargus*) between Europe and Sub Saharan Africa. I reconstructed the position of those birds from the West Europe breeding sites, migrating in Autumn to the westernmost sites of Africa through the western migration route. The reconstructed positions and the manifold show a slight curvature (Fig. 6a). Trieweiller et al. (2014) discussed the characteristics of this route that explain the curvature, i.e. a short crossing distance across Mediterranean Sea (the Gibraltar Strait), and increased amount of potential foraging habitat (as compared with other migration flyways). Another important organism performing migrations in N Africa is the desert locust, *Schistocerca gregaria*, (Draper 1980). Swarms located in the West Sub Saharan Africa summer breeding area cross to Asia and return, while swarms in west Africa appear to migrate within the continent through various routes (based on Rainey 1963, cited in Draper 1980). The manifold for the animals migrating from Asia (Fig. 6b) captures the general pattern of migration, while the curvature upwards reflects a potential wide area of crossing over the Red Sea. These two examples introduce a second type of application of the manifolds: one could study the effect of distance-dependent effects of migrants on organisms living on or near that route: for instance, as locusts (or birds) migrate, they may stop to graze (or hunt), thereby impacting primary producers (or consumers) living along the routes. The influence of such migrants on the spatial distribution of resources across the landscape should be distance-dependent and may be better explained with reference to the curvilinear coordinate system arising from the pattern of migration.

Manifolds may have applications in conservation, for instance in the design of marine protected areas. For example, Marancik et al. (2012) presented data on spatial distribution of larval stages of grouper, an economically important fish, distributed across the Gulf of Mexico. In the northern sector of the gulf, early stage larvae concentrate along the shelf break. The manifold (Fig 6c) follows closely the patterns of distribution and the shelf break. Grouper are overfished (Marancik et al. 2012) and the approximation of manifolds could be used to establish marine protected areas (MPAs) if such areas are mapped with respect to manifold coordinates. The issue of MPAs is

highlighted in the study of Planes et al. (2009) who mapped the position of anemones inhabited by the orange clownfish (*Amphiprion percula*) around Kimbe Island, in Papua New Guinea. The manifold (Fig. 6d) approximated from the spatial position of such anemones follows roughly the coastline curvature of Kimbe Island. Distances among anemones on the manifold may be used to study the process of dispersal and connectivity at the spatial scale covered of the whole coastline.

Regional manifolds may be useful to test alternative EGDs explaining community connectivity, such as those inquiline metacommunities existing in the pitcher plant *Sarracenia purpurea* (Buckley et al. 2003, Milner and Kneitel 2005). Pitcher plants accumulate water in their (cup shaped) leaves and hence support local communities of invertebrates and protozoans. Buckley et al. (2003) studied latitudinal gradients in species richness in those meta-communities across the North American continent. In the USA, the distribution of pitcher plants is restricted to the eastern seaboard; I used spatial positions of those populations studied by Buckley et al. (2003) and those in southern Canada to fit a manifold that may explain patterns of connectivity. Other populations were not used because pitcher plants show a widespread distribution across Canada and 1D manifolds will not be useful to explain connectivity. The manifold (Fig. 6e) captures quite well the curvature of the geographic distribution of pitcher plants. Note however that some projections lie outside the range of distribution; this is likely due to the small number of data points used to approximate the manifold as well as the strong curvature of the distribution. This issue may be solved by approaching the manifold to a larger number of local populations in the map. Notwithstanding the issue, the connectivity among plants will be explained mostly by the distances along the manifold (originating axis z_1, see example in figure 5b).

6. Discussion

Perhaps the main contribution of spatial ecology, meta-population theory and landscape ecology have been in highlighting the ecological importance of spatial distances between interacting organisms, or between different types of resources or habitable patches. Such contributions have challenged the assumptions of closed and well-mixed populations. This article highlights the curved nature of the spaces where many ecological processes occur (by extension distances among organisms are not necessary linear). Hence, this is a contribution, along with other approaches (e.g. fractal distances: Edwards et al. 2007, oceanographic distance: White et al. 2010, minimum cost

distances: Geenberg et al. 2011), to make Euclidean or geographical distances testable assumptions in studies of patterns of species distribution.

The main idea proposed here is to view migration as if organisms were moving along curvilinear geometrical objects (smooth manifolds), shaped by the interactions between behaviour and landscape. In this view, the curvilinear pathways are the "shortest realised paths" between habitat patches, an idea that is consistent with the concept of "minimum cost corridors/surfaces" used in landscape ecology (Estrada-Peña 2003, Geenberg et al. 2011). Such "shortest paths", defined on the manifolds, become segments of the manifold geodesics. In physics, manifold geodesics are interpreted as the pathways of minimal energy (e.g. the circular trajectory of a passive particle captured in an eddy). This will be also the case if organisms drift with the wind or water currents. However, migration/dispersal pathways are usually shaped by behavioural responses to the elements of the landscape (food, predators, topography); in many instances, such pathways result from a process of optimization, driven by selective forces such as the necessity to minimise energy costs and mortality risks (Alerstam et al. 2003). In such scenario, the manifold geodesics constitute pathways of maximum fitness shaped by natural selection.

The introduction of manifolds and the definition of geodesics led to the formulation of a procedure to test of alternative EGDs as metrics of distance-dependent phenomena (e.g. connectivity, species interactions). Such metrics enables one to formulate mechanistic models that result in linear statistical models. The models are mechanistic at the ecological level because distinct manifold geometries (when constructed as in Fig. 3) connect different patterns of animal behaviour, biotic and abiotic factors (e.g. landscape topography, the quality of the habitat matrix) with patterns of species distributions. The models are linear thanks to the coordinate systems defined on curved spaces. The models can be applied to spatial position of habitat patches or organisms depending on the question; in particular this application will become useful as data on track annotation (i.e. records of spatial positions of migrating organisms while *en route*) become available (e.g. see Mandel et al. 2011). In this case, each individual path becomes a replicate unit and the manifold is the estimated average path followed by migrants.

The geometrical view to transport and dispersal will be useful in the field of meta-population, meta-community ecology (Connolly et al. 1999, Armsworth 2002, Cowen 2006, Holyoak et al. 2005). There are a number of potential applications related to

climate change, habitat fragmentation, invasive species and the transmission of diseases (Kokko and López-Sepulcre 2008, Holyoak and Heath 2016). For instance, in response to habitat fragmentation and overfishing, the design of networks of marine protected areas consider distances as key criteria to ensure connectivity (Costello et al. 2010, Gaines et al. 2010) and such distances arise from the geometry of the larval dispersal plume. Another area of application is the dynamics of invasion, considered one of the most important threats to biodiversity (Bellard et al. 2013, Seebens et al. 2013): here manifolds can help in identifying the paths to invasion (as shown in the idealised example discussed in section 3).

The geometrical view to animal movement proposed here should have important applications in landscape ecology. For instance, one can test for different metrics of distance dependent interactions, occurring e.g. while organisms are *en route* (i.e. as migrants interact with organisms of the habitat matrix or as an invading species undergoes range expansion). Migrants (e.g. locusts: Simpson et al. 1999) may consume food and modify the biotic components of the landscape. In addition, some migrating organisms (e.g. geese: Giroux et al. 2012) subsidise populations located along or at the end of migrating pathways. Hence, manifold approaches could improve our understanding of the role of allochthonous subsidization in food web structure. Migrating birds are vectors of viruses that can infect humans (Lam et al. 2012); Lam et al (2012) showed that phylogeographic models incorporating information on the migration routes improved the fit of sequence data on influenza virus, as compared to models assuming homogeneous or random rates of gene flow among localities. Manifolds can help the development of such models with the incorporation of curvilinear geometry of dispersal pathways in e.g. isolation by distance models.

In the types of applications discussed above, the geometric approach proposed here enables the formulation of testable mechanistic models. For instance, in the marine environment, dispersal patterns are driven by hydrodynamic features and larval vertical migration; this is shown in the example of Ayata et al. (2010) where different manifolds were approximated by the NLPCA. There are in addition other similar examples based on numerical modelling (Marta-Almeida et al. 2006, Paris et al. 2007, Robins et al. 2013) as well as field studies on larval behaviour and distribution (Forward and Tankersley 2001, Queiroga and Blanton 2004, Shanks et al. 2005). Likewise, insect migrations are driven by behaviour (e.g. avoid flying over sea: Shashar et al. 2005) and atmospheric processes (Drake and Farrow 1988, Chapman et al. 2010). Bird migratory

routes are also shaped by climatic conditions, landscape topography as well as the distribution of foraging habitats (Mandel et al. 2011, Trierweiller et al. 2014). In all these scenarios it is likely that hypothesis testing favours a curvilinear EGD. In other scenarios, hypothesis testing may favour Euclidean or geographic distances as EGD (e.g. locust swarms in West Africa: Draper 1980; inquiline communities of pitcher plants in North Canada: Buckley et al. 2003).

There are cases, where trajectories are not smooth; in some of those cases, trajectories may be better approximated as fractal objects (e.g. Schneider 1994, Cap 5); connectivity on rivers and streams can be modelled through different types of dendritic network approaches (Peterson et al. 2013, Melles et al. 2015). In cases where trajectories are smooth but show self-crossings, the geometric approach proposed here is not valid because it is not possible to produce a one to one mapping of each point of the manifold (e.g. if the trajectory is looking like "∞" there is are two points showing the same coordinates).

It is likely that whether curvilinear EGDs are better at explaining animal movement depends on the scale of observation. For instance, in inquiline meta-communities in pitcher plants, Milner and Kneitel (2005) defined three spatial scales of connectivity: among leaves of the same plant, among plants from the same local (plant) population, and among plant populations distributed across the North American continent. Curvilinear dispersal may best represent patterns at the largest of the three spatial scales. Dispersal among leaves of the same plant may be viewed as occurring in 3D Euclidean space and hence better explained by Euclidean distances. More generally, at small scales, animal movements have been abstracted as fractal objects or correlated random walks arising from the landscape topography and behaviour (Crist et al. 1992; Edwards et al. 2007). Obviously, the concept of "small scale" will vary with body size and movement capabilities because animals as small as heterotrophic protozoans and as large as birds exhibit such behaviours (Bartumeus et al. 2003; Edwards et al. 2007). The fact that movement patterns and behaviour may be scale-dependent is consistent with the notion that smooth manifolds can be considered locally flat but regionally curved.

At least three statistical issues should be kept in mind at the time of using manifold approximation. First, there are different methods (Gorban et al. 2008) and my goal was not to promote a specific one, but to explore the potential of such techniques. Second, manifold approximation techniques are based on iterative methods that are sensitive to the initial values and they can be stuck in local minima (Gorban et al. 2008). A key

point is to avoid model overfitting, which results in overly complex manifolds characterised by an excessive number of loops. For NLPCA, model complexity is controlled by a penalty term and Scholz (2012) provides a procedure for model validation using such term. The manifold, as any other statistic, is sensitive to the number and dispersion of the observations across the landscape; hence, a better approximation of the manifold will be achieved with large samples. Third, at the stage of model selection/hypothesis testing, the usual issues arising from assumptions about distribution of residual errors or spatial autocorrelation (Dorman et al. 2007; Kissling and Carl 2008, Zuur et al. 2009) will apply. The geometric approach proposed here addresses the problem of model specification, which is the most important issue around model fitting (Faraway 2014); this is achieved through the transformation of spatial coordinates defined on a Euclidean (or geographical) space into coordinates on a curvilinear manifold.

Overall, a geometrical view to animal movement, as occurring along smooth manifolds, could further help us to explain how such distance-dependent processes drive patterns of species distribution. This approach enables the test of alternative definitions of distance, either Euclidean, geographic or curvilinear at least in three areas of ecology. Within meta-population (and community) ecology one could use them to evaluate the consequence of connectivity on ecological patterns; in this case, the manifold approach enables the correct calculation of distances between source and receiving habitat patches. Within landscape ecology the introduction of manifolds enables the correct calculation of distances along the path of migration in order to study e.g. the effects of migrants on organisms living along the routes. The development of numerical models of larval transport and the procedures of animal track annotation along the migration routes should make manifolds a very useful tool in understanding the role of dispersal and migration in ecological patterns.

Acknowledgments

I appreciate comments and critiques of a draft version by S. Jenkins; in addition, this article has improved thanks to the comments made by the reviewers. I also acknowledge support by SEACAMS2 (Sustainable Expansion of the Applied Coastal and Marine Sectors 2) (Operation number 80860) part funded by the European Regional Development Fund (ERDF) from the Welsh European Funding Office (WEFO).

I do not have any conflict of interest.

References

Adriaensen, F. et al. 2003. The application of 'least-cost' modelling as a functional landscape model. – Landsc. Urb. Plan. 64:233–247.

Alberto, F. et al. 2011. Isolation by oceanographic distance explains genetic structure for *Macrocystis pyrifera* in the Santa Barbara Channel. – Molec. Ecol. 20:2543–2554.

Alerstan T. 2003. Long-distance migration: evolution and determinants. – Oikos 103:247–260.

Armsworth, P.R. 2002. Recruitment limitation, population regulation, and larval connectivity in reef fish metapopulations. – Ecology 83:1092–1104.

Ayata, S-D. et al. 2010. How does the connectivity between populations mediate range limits of marine invertebrates? A case study of larval dispersal between the Bay of Biscay and the English Channel (North-East Atlantic). – Progr. Oceanogr. 87:18–36.

Bartumeus, F. et al. 2003. Helical Levy walks: Adjusting searching statistics to resource availability in microzooplankton. – Proc. Natl. Acad. Sci. USA 100:12771–12775.

Bellard, C. et al. 2013. Will climate change promote future invasions? – Glob. Change. Biol. 19:3740–3748.

Boedeltje, G. et al. 2003. Plant dispersal in a lowland stream in relation to occurrence and three specific life-history traits of the species in the species pool. – J. Ecol. 91:855-866.

Both, C. 2010. Flexibility of timing of avian migration to climate change masked by environmental constraints en route. – Current Biology 20:243–248.

Buckley, H.L. et al. 2003. Reverse latitudinal trends in species richness of pitcher-plant food webs. – Ecol. Lett. 6:825–829.

Butler, M.I. et al. 2011. Behavior constrains the dispersal of long-lived spiny lobster larvae. – Mar. Ecol. Prog. Ser. 422:223–237.

Carlton, J.T. and Hodder, J. 1995. Biogeography and dispersal of coastal marine organisms: experimental studies on a replica of a 16th-century sailing vessel. – Mar. Biol. 121:721–730.

Chapman, J.W. et al. 2010. Flight orientation behaviors promote optimal migration trajectories in high-flying insects. – Science 327:682–685.

Connolly, S.R. et al. 2001. A latitudinal gradient in recruitment of intertidal invertebrates in the northeast Pacific Ocean. – Ecology 82:1799–1813.

Cowen, R.K. et al. 2006. Scaling of connectivity in marine populations. – Science 311:522–527.

Costello, C. et al. 2010. The value of spatial information in MPA network design. – Proc. Natl. Acad. Sci. USA 107:18294–18299.

Crist, T.O. et al. 1992. Animal movement in heterogeneous landscapes: An experiment with *Eleodes* beetles in shortgrass prairie. – Funct. Ecol. 6:536–544.

D'Aloia, C. et al. 2015. Patterns, causes, and consequences of marine larval dispersal. – Proc. Natl. Acad. Sci. USA 112:13940–13945.

Diekmann, U. et al. 2000. The geometry of ecological interactions. Simplifying spatial complexity. – Cambridge.

Dormann, C.F. et al. 2007. Methods to account for spatial autocorrelation in the analysis of species distributional data: a review. – Ecography 30:609–628.

Drake, V.A. and Farrow, R.A. 1988. The influence of atmospheric structure and motions on insect migration. – Annu. Rev. Entomol. 33:183–210.

Draper, J. 1980. The direction of desert locust migration. – J. Anim. Ecol. 49:959–974.

Edwards, A.M. et al. 2007. Revisiting Levy flight search patterns of wandering albatrosses, bumblebees and deer. – Nature 449:1044–1048.

Einbeck, J. et al. 2010. Data compression and regression through local principal curves and surfaces. – Int. J. Neur. Syst. 20:177-192.

Estrada-Peña, A. 2003. The relationships between habitat topology, critical scales of connectivity and tick abundance *Ixodes ricinus* in a heterogeneous landscape in northern Spain. – Ecography 26:661–671.

Failmezger, H. et al. 2013. Semi-automated 3D leaf reconstruction and analysis of *Trichome* patterning from light microscopic images. – PLOS Computational Biology 9:e1003029.

Faraway, J. 2014. Linear models with R. – CRC Press.

Forward, R.B. and Tankersley, R.A. 2001. Selective tidal-stream transport of marine animals. – Oceanogr. Mar. Biol. Annu. Rev. 39:305–353.

Gaines, S.D. et al. 2010. Designing marine reserve networks for both conservation and fisheries management. – Proc. Natl. Acad. Sci. USA 107:18286.

Gorban, A.N. et al. 2008. Principal manifolds for data visualisation and dimension reduction. – Springer.

Gorban A.N. and Zinovyev, A. 2008. Elastic maps and nets for approximating principal manifolds and their application to microarray data visualization. – In: Gorban, A.N.,

Kégl, B., Wunsch, D.C., Zinovyev, A. (eds.) Principal manifolds for data visualisation and dimension reduction. Springer. pp 97–131.

Gorban A.N. and Zinovyev, A. 2010. Principal manifolds and graphs in practice: from molecular biology to dynamical systems. – Int. J. Neur. Syst. 20:219–232.

Greenberg, J. A. et al. 2011. Least cost distance analysis for spatial interpolation. – Comput. Geosci. 37:272–276.

Grinfeld, P. 2013. Introduction to tensor analysis and the calculus of moving surfaces. – Springer.

Hanski I. 1999. Metapopulation Ecology. – Oxford.

Holyoak, M. and Heath, S.K. 2016. The integration of climate change, spatial dynamics, and habitat fragmentation: A conceptual overview. – Integr. Zool. 11:40–59.

Holyoak, M., et al. 2005. Metacommunities. Spatial dynamics and ecological communities. – Univ. Chicago Press.

Hsieh, W.W. et al. 2006. Nonlinear atmospheric teleconnections. – Geoph. Res. Lett. 33:L07714.

Isaaks, E. and Srivastava, R. 1989. – Applied Geostatistics. Oxford.

Kissling, W.D. and Carl G. 2008. Spatial autocorrelation and the selection of simultaneous autoregressive models. – Global. Ecol .Biogeog. 17:59–71.

Kokko, H. and López-Sepulcre, A. 2006. From individual dispersal to species ranges: Perspectives for a changing world. – Science 313:789–791.

Lam, T.T-Y. et al. 2012. Migratory flyway and geographical distance are barriers to the gene flow of influenza virus among North American birds. – Ecol. Lett. 15:24–33.

Legendre, P, and Legendre, L. 1998. Numerical ecology. – Springer.

Levin, A.S. 1992. The problem of pattern and scale in ecology. – Ecology 73:1943–1967.

Marancik, K.E. et al. 2012. Spatial and temporal distribution of grouper larvae (Serranidae: Epinephelinae: Epinephelini) in the Gulf of Mexico and Straits of Florida. – Fish. Bull. 110, 1–20.

Marta-Almeida, M. et al. 2006. Influence of vertical migration pattern on retention of crab larvae in a seasonal upwelling system. – Mar. Ecol. Prog. Ser. 307:1–19.

Massol, F. et al. 2011. Linking community and ecosystem dynamics through spatial ecology. – Ecol. Lett. 14:313–323.

Melles, S.J. et al. 2015. Potential spread of Great Lakes fishes given climate change and proposed dams: an approach using circuit theory to evaluate invasion risk. – Landscape Ecology 30:919–935.

McIntyre, N.E. and Wiens J.A. 1999. How does habitat patch size affect animal movement? An experiment with darkling beetles. – Ecology 80:2261–2270.

Millner, T. and Kneitel, J.M. 2005. Inquiline communities in pitcher plants as a prototypical metacommunity. – In: Holyoak, M., Leibold, M.A. and Holt, R.D.(eds). Metacommunities. Spatial Dynamics and Ecological Communities. Univ. Chicago Press. pp. 122–145.

Moritz, C. et al. 2013. Disentangling the role of connectivity, environmental filtering, and spatial structure on metacommunity dynamics. – Oikos 122:1401–1410.

Paris, C.B. et al. 2007. Surfing, spinning, or diving from reef to reef: effects on population connectivity. – Mar. Ecol. Prog. Ser. 347:285–300.

Pearson, K. 1901. On lines and planes of closest fit to systems of points in space. – Phil. Magazine 2:559–572.

Planes, S. et al. 2009. Larval dispersal connects fish populations in a network of marine protected areas. – Proc. Natl. Acad. Sci. USA 106:5693–5697.

Plant, R. 2012. Spatial data analysis in ecology and agriculture using R. – CRC Press.

Queiroga, H. and Blanton, J. 2004. Interactions between behaviour and physical forcing in the control of horizontal transport of decapod crustacean larvae. – Adv. Mar. Biol. 47:107–214.

Rietkerk, M. and van de Koppel. J. 2008. Regular pattern formation in real ecosystems. – Trends. Ecol. Evol. 23:169–175.

Robins, P.E. et al. 2013. Physical and biological controls on larval dispersal and connectivity in a highly energetic shelf sea. – Limnol. Oceanogr. 58:505–524.

Scholz, M. et al. 2008. Nonlinear principal component analysis: Neural network models and applications. – In: Gorban, A.N., Kégl, B., Wunsch, D.C., Zinovyev, A. (eds.) Principal manifolds for data visualisation and dimension reduction. Springer. pp. 45–68.

Schneider, D.C. 1994. Quantitative ecology. – Academic Press. London.

Seebens, H. et al. 2013. The risk of marine bioinvasion caused by global shipping. – Ecol. Lett. 16:782–790.

Shanks, A. and Brink, L. 2005. Upwelling, downwelling, and cross-shelf transport of bivalve larvae: test of a hypothesis. – Mar. Ecol. Prog. Ser. 302:1–12.

Shashar, N. et al. 2005. Migrating locusts can detect polarized reflections to avoid flying over the sea. – Biol. Lett. 1:472–475.

Stamps, J.A. et al. (1987). The effects of edge permeability and habitat geometry on emigration from patches of habitat. – Am. Nat. 129:533–552.

Tilman, D. and Kareiva, P. 1997. Spatial ecology. The role of space in population dynamics and interspecific interactions. – Princeton.

Trierweiler, C. et al. 2014. Migratory connectivity and population-specific migration routes in a long-distance migratory bird. – Proc. R. Soc. B: Biol. Sci. 281:20132897.

White, C. et al. 2010. Ocean currents help explain population genetic structure. – Proc. R. Soc. B: Biol. Sci. 277:1685–1694.

Wiens, J.A. 1989. Spatial scaling in ecology. – Funct. Ecol. 3:385–397.

Wiens, J.A. and Bachelet, D. 2010. Matching the multiple scales of conservation with the multiple scales of climate change. – Cons. Biol. 24:51–62.

Zuur, A. et al. 2009. Mixed effect models and extensions in ecology with R. – Springer.

Figure Legends

Figure 1. Idealised trajectories during larval transport associated to an eddie system. (1) Populations located at sea bottom habitats (red patches) release larvae (trajectories as blue dots) that ascend and are captured by an eddie (light blue disc with rotating arrow).. Larval pathways are constrained by behaviour (upwards migration) and the eddie circulation. (2) The collection of pathways delineates a region of a 2D smooth manifold (in green).

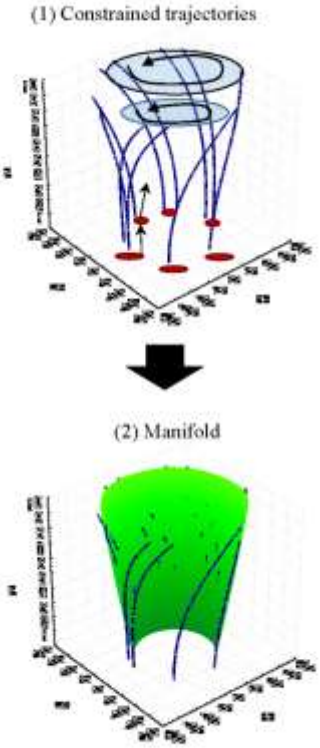

Figure 2. Understanding curvature and geodesics. The panels depict a series of points and trajectories on two manifolds (a plane and a paraboloid) placed on an 3D Euclidean space where a Cartesian coordinate system (x,y,z) is defined. For the case of the paraboloid there is tangent plane (ellipses) i.e. touching the point in question, but not crossing to the other side. At any point (black circles), three vectors, perpendicular to each other, are defined: the *tangent vector* (T blue), touching the point in the direction of the path; the *tangential normal vector* (n, red), perpendicular to the tangent vector but in the plane; the *surface normal* (N, thick black arrows), perpendicular to the plane pointing away from the manifold. **(A)** Displacements along a straight line (blue) results in that all vectors remain parallel; displacement along a curve makes (red) makes the tangent vector change direction; in consequence the tangential normal also changes direction. **(B)** Displacement on a curvilinear manifold along a geodesic path (blue). The tangent vector changes direction showing that the path is curved. The path bends on the 3D space, defining a vector, proportional to the surface normal and characterised by the normal curvature (k_n). Notice that the surface normal changes direction. However, the tangential normal remains parallel along the path: this property is shared with the straight path in (A): they are both geodesics on their respective manifolds (the plane and the paraboloid). **(C)** Non geodesic path (red): The path now curves on the manifold, defining a vector (proportional to the tangential normal) characterised by the geodesic curvature (k_g). Notice that now the tangential normal changes direction as in the curved line in (A). Because of the extra curvature ($k_g > 0$) distances between two points along any non-geodesic curve will be longer than those defined by the geodesics.

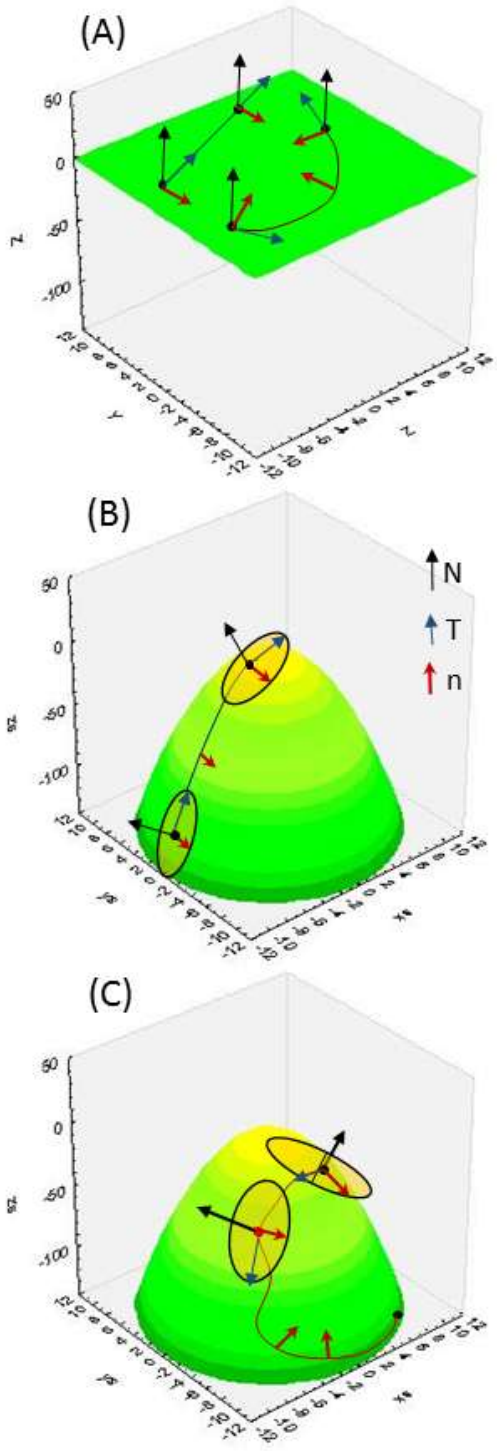

Figure 3. Summary of modelling scenarios (a,b) and abundance distributions (c,d). (a) Idealised bay with 350 patches colonised by a species from the NE starting with habitat patch h_1; habitat patches (h_i) are indicated as black dots. Thin arrows: direction of natural transport of marine species along the bay; coastal waters are separated from offshore water by a front. Thick arrows: natural migration pathways followed by terrestrial species. Hatched arrows: maritime transport (only two paths are indicated). (b) Three idealised parabolic bays, with different curvature parameter (*σ-value)*. Modelled average abundance distributions are shown in (c) for the natural transport and (d) for maritime transport; in both cases for $\sigma = 0.5$ and decay constant =0.1 (for simplicity, numbers in z axis were removed, but are shown in Figs. S2 and S4).

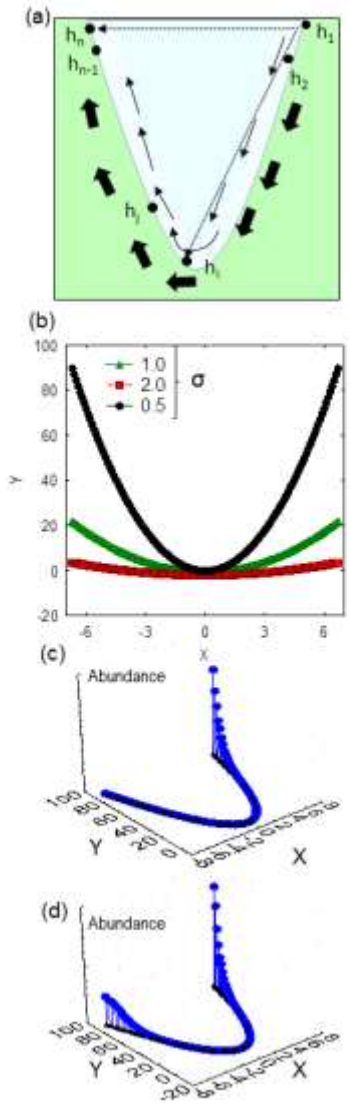

Figure 4. Summary of results of model simulations for three idealised bays varying in shoreline curvature for a decay constant β=0.2 (see supplementary figures for other β-values). Left panels: average of simulated abundance patterns along the bays. Right panel: ΔAIC calculated for each simulation, comparing true model (natural transport mechanism) modelled as a parabola (P) vs maritime transport mechanism with abundance depending on x-coordinates only (1D), on $x+y$ coordinates (2D) and on Euclidean distance (Euc) from the habitat patch (see Table 1 for equations). Positive ΔAIC: AIC of the alternative model is higher than that of the true model.

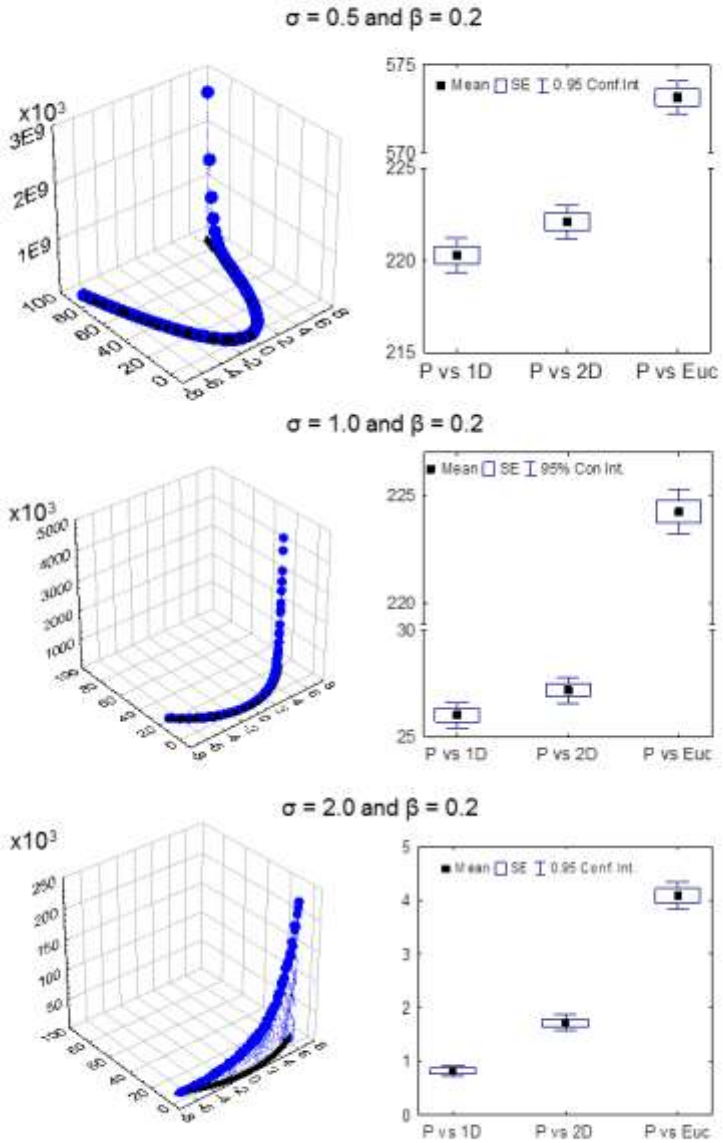

Figure 5. Examples of manifold approximation based on modelled larval pathways and dispersal kernels. (a) Manifold approximated from dispersal pathways obtained from numerical models of transport of spiny lobster larvae in the Yucatan Bay (source: Butler et al. (2011). (b) New coordinate system derived from the projections of the dispersal plume on the manifold and the distances between the projection and each point of the plume. Blue symbols: spatial position of selected sites of larval release. (c-d) Manifolds approximated from dispersal plume obtained by modelling transport under different types of larval behaviour in Bay of Biscay (source: Ayata et al. 2010).

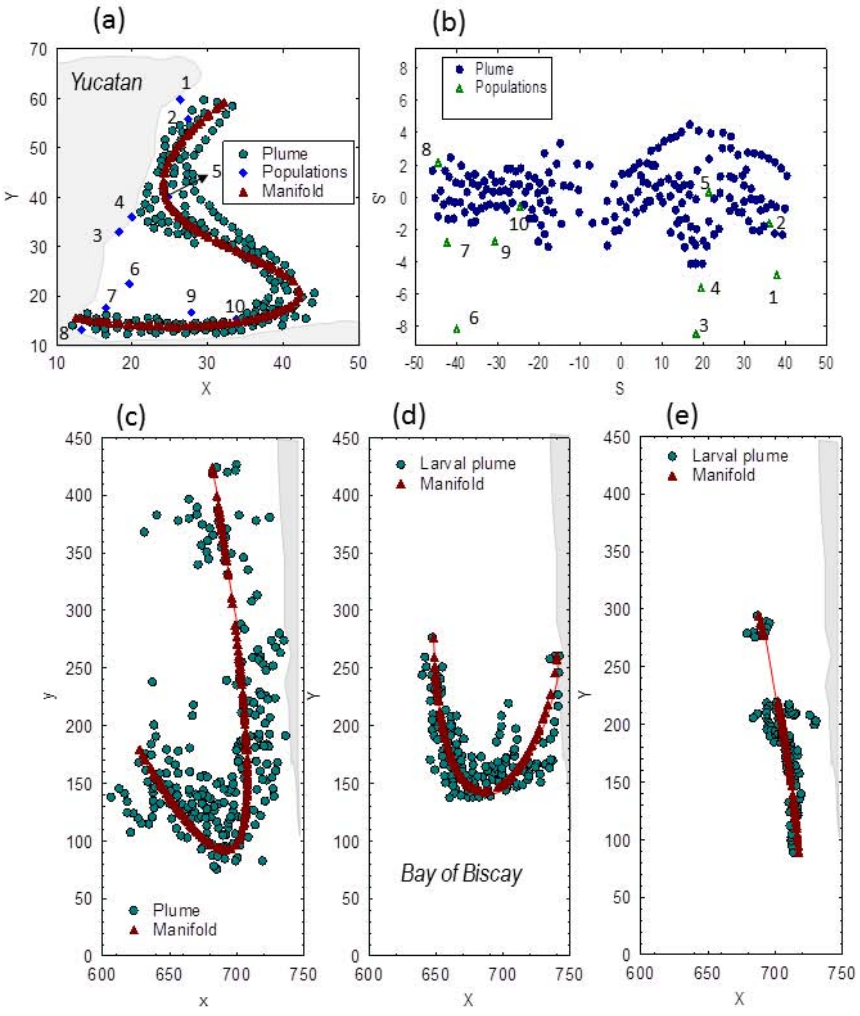

Figure 6. Examples of manifold approximation based on actual data. (a) migration pathways of Montagu's harriers crossing from Europe of sub-Saharan Africa, (b) desert locus swarm distribution and migration, (c) patterns of abundance of fish larvae, (d) spatial distribution of anemones and (e) spatial distribution of plants. Sources are as follows (a) Trierweiller et al. (2014); (b) Draper (1980); (c) Mrancick et al. (2012) (d) Planes et al. (2009), (e) Buckley (2003).

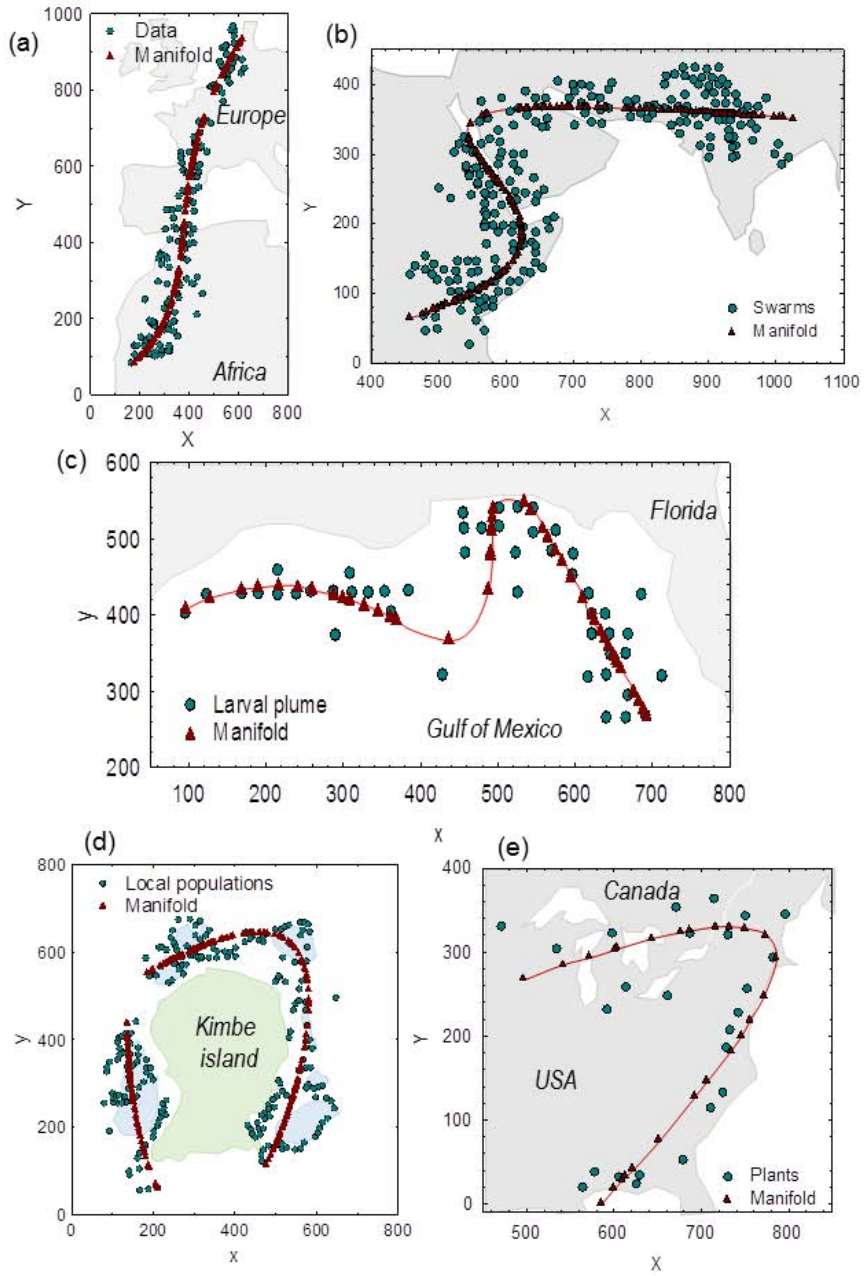

Table Legends

Table 1. Models used to test different hypothetical scenarios of larval dispersal, the associated manifold, coordinate systems and predictor variables.

Models	Manifold	Coordinate system	Equation
True model (Parabola)	1D Parabola	Arc-length	$\ln(N) = a + bz$, z= arc length of parabola
1D	1D flat space	Cartesian	$\ln(N) = a + b \cdot x$
2D	2D flat space	Cartesian	$\ln(N) = a + b \cdot x + c \cdot y$
Euclidean	2D flat space	Cartesian	$\ln(N) = a + b \cdot h$, $h = (x'^2 + y'^2)^{1/2}$ $x' = x + 6.7$, $y' = y(x) - y(x')$

Table 2. Average ΔAIC between models using Cartesian system vs the true model (Z: based on Arc length coordinates) for nine combinations of distance decaying (β) and coastline shapes (σ) parameters.

	2D vs Z			1D longshore vs Z			Euclidean vs Z		
$\sigma \rightarrow$	0.5	1	2	0.5	1	2	0.5	1	2
β = 0.1	97.2	7.8	1.1	95.7	6.8	0.2	391.8	97.3	1.0
β = 0.2	222.1	27.2	1.7	220.3	26.0	0.8	573.1	224.2	4.1
β = 0.3	317.5	54.0	2.6	315.6	52.7	1.5	682.7	317.7	8.4