ICES Journal of Marine Science, 2022, **79**, 609–632
DOI: 10.1093/icesjms/fsac021
Advance access publication date: 2 March 2022
Original Article

Quantitative uncertainty estimation in biophysical models of fish larval connectivity in the Florida Keys

Romain Chaput [1,2,*], Pierre Sochala [3,4], Philippe Miron [5], Vassiliki H. Kourafalou [1] and Mohamed Iskandarani [1]

[1]Department of Ocean Sciences, Rosenstiel School of Marine and Atmospheric Science, University of Miami, Miami, FL 33149, USA
[2]School of Biological Sciences, Victoria University of Wellington, Wellington 6012, New Zealand
[3]BRGM , 3 Avenue Claude Guillemin, 45060 Orleans, France
[4]CEA, DAM, DIF, F-91297 Arpajon, France
[5]Center for Ocean-Atmospheric Prediction Studies, Florida State University, Tallahassee, FL 32310, USA

[*] *Corresponding author: tel no:+64 4 463 5339;* e-mail: rchaput@rsmas.miami.edu.

The impacts of seven uncertain biological parameters on simulated larval connectivity in the Florida Keys are investigated using Polynomial chaos surrogates. These parameters describe biological traits and behaviours—such as mortality, swimming abilities, and orientation—and modulate larval settlement as well as dispersal forecasts. However, these parameters are poorly constrained by observations and vary naturally between individual larvae. The present investigation characterizes these input uncertainties with probability density functions informed by previous studies of *Abudefduf saxatilis*. The parametric domain is sampled *via* ensemble calculations, then a polynomial-based surrogate is built to explicitly approximate the dependence of the model outputs on the uncertain model inputs, which enables a robust statistical analysis of uncertainties. This approach allows the computation of probabilistic dispersal kernels that are further analyzed to understand the impact of the parameter uncertainties. We find that the biological input parameters influence the connectivity differently depending on dispersal distance and release location. The global sensitivity analysis shows that the interactions between detection distance threshold, orientation ontogeny, and orientation accuracy, are the dominant contributors to the uncertainty in settlement abundance in the Florida Keys. Uncertainties in swimming speed and mortality, on the other hand, seem to contribute little to dispersal uncertainty.

Keywords: biophysical model, connectivity, coral reef, dispersal, fish larvae, orientation, Polynomial chaos analysis, uncertainty quantification.

Introduction

Biophysical models of dispersal are increasingly used to estimate population connectivity of many sessile and benthic marine species (Swearer *et al.*, 2019; Andrello *et al.*, 2013; Burgess *et al.*, 2014). Populations of these species are mostly connected by the exchange of larvae that are transported in the open ocean during the dispersal phase (Cowen and Sponaugle, 2009). Biophysical models are designed to simulate this transport and the biophysical interactions happening during dispersal (Paris *et al.*, 2005; Ayata *et al.*, 2009; Andrello *et al.*, 2013; Almany *et al.*, 2017). However, the utility of these biophysical models in informing science and resource management—for example, determining which biological traits are most relevant to connectivity (Staaterman *et al.*, 2012), helping the placement of Marine Protected Areas (Andrello *et al.*, 2015; Claro *et al.*, 2018; Kough *et al.*, 2018), or advising sustainable fisheries management (García-García *et al.*, 2016)—is contingent on the confidence in the models' output. This confidence can be enhanced by estimating the uncertainties in the biophysical model outputs caused by uncertainties in the model formulation and poorly constrained input parameters. The study, herein, presents a novel application of surrogate-based Uncertainty Quantification methodology to estimate the impact of uncertain biological parameters on simulated fish larval connectivity in the Florida Keys. The quantification of the impact of these uncertainties aims at evaluating the reliability of connectivity estimates.

Biophysical models simulate larval exchange and transport in the open ocean, first, by replicating oceanic conditions as accurately as possible, and second, by representing larvae as particles endowed with certain attributes that model, as faithfully as possible, known larval behaviours (Swearer *et al.*, 2019). The primary goal of biophysical connectivity models is in reproducing the journey of successful larvae, the ones that find their way to a reef and settle. Indeed, only larvae that survive the pelagic phase take part in shaping populations connectivity. If most of the larvae were to survive the larval phase, one could use the average values of the settling individuals to represent the common biological and behavioural traits of the larvae modelled. However, very few larvae survive the pelagic phase and successfully settle on the reefs (Houde, 2002); the ones that do survive potentially have traits very different from the others. Indeed, studies have shown that successful larvae have specific traits such as faster growth (Shulzitski *et al.*, 2015; Goldstein and Sponaugle, 2020) and/or better orientation abilities (Staaterman *et al.*, 2012). Therefore, dispersal models should not use only averaged values of larval traits as they would misrepresent the larvae fraction that successfully participate in population connectivity. One way to mitigate this problem is to characterize the uncertainty in the biological input parameters by their observed distributions and consider runs of the biophysical model as stochastic realizations of a probabilistic process (Leis, 2007). Such approach would simulate the dispersal for the population of larvae rather than

Received: September 16, 2021. **Revised:** January 23, 2022. **Accepted:** January 24, 2022

a subset, allowing an evaluation of the reliability of the model and a better understanding of the impact of biological traits on population connectivity.

The accuracy and reliability of connectivity estimates hinge on the biophysical model's ability to correctly represent mathematically the interaction between oceanic transport and larvae behaviours (Staaterman and Paris, 2013). Higher reliability can be achieved by: (i) using high resolution oceanographic models, (ii) choosing experimentally validated biological and behavioural traits, and (iii) quantifying uncertainties (Ådlandsvik *et al.*, 2009). Uncertainties in forecast model, such as biophysical models of larval fish, can be categorized into two classes: the *model uncertainties* caused by the abstraction of the true processes (physical or biological) into mathematical equations and the *input data uncertainties* due to the empirical parameters involved in the model itself. The focus of this study is the characterization of this second type of uncertainties. Uncertainties in biophysical models originate from the oceanographic and biological model components, but the impacts of uncertain biological parameters are far less documented and quantified than those of oceanographic ones (see Clancy *et al.* (2010); Hu *et al.* (2012); and Mattern *et al.* (2013) for a few examples). There are a lot of uncertainties in the biological part of the model since it aims at reproducing an inherently variable and adaptive system. In addition, some of the active biological behaviours recently implemented in biophysical models (Staaterman and Paris, 2013; Simpson *et al.*, 2013) rely on parameters, such as onset of orientation, individual accuracy, and detection distance of the cues, for which precise measurements (theoretical or empirical) are lacking (Leis, 2007), adding to the uncertainties.

The objective of this study is to evaluate the reliability of a complex biophysical model that includes a larval orientation component. This is done by following four classical Uncertainty Quantification analysis steps (Iskandarani *et al.*, 2016). First, the model input uncertainties are identified and quantified. The present study focuses exclusively on biological parameters that simulate larvae behaviour known to impact the probability of settlement (Leis, 2006; Staaterman *et al.*, 2012). A total of seven uncertain inputs, related to the larvae biological and behavioural models, are taken into account. The characterization of the input uncertainties (i.e. specifying the parameter ranges and distributions) were informed by empirical information collected on *Abudefduf saxatilis*, a common species of Pomacentridae with documented ecology and larval traits (Fishelson, 1970; Foster, 1987; Prappas *et al.*, 1991; McAlary and McFarland, 1993; Aishuth *et al.*, 1998; Hogan and Mora, 2005). Second, the uncertainties are forward propagated into a biophysical model, the Connectivity Modelling System (CMS, Paris *et al.* (2013b)). We perform an ensemble of simulations in order to sample the parametric domain and compute the corresponding outputs. Third, the ensemble data are used to build and validate surrogate models of the dispersal kernels. Polynomial chaos (PC) surrogates are implemented, non-intrusively, to explicitly approximate the dependence of the model outputs on the uncertain model inputs. Fourth, once validated, the surrogates replace the original model for the statistical analysis. This approach enables the generation of a probabilistic connectivity estimate whose properties can be readily analyzed to determine the dominant contributors to the output uncertainties.

The paper is structured as follows. Material and methods introduce the general setting of this work including the biophysical model of dispersal, the oceanic conditions with the associated Lagrangian Coherent Structures (LCS), the biological and behavioural models, the parametric domain, and the quantities of interest. The section Uncertainty propagation details the probabilistic framework associated with the biophysical model of dispersal. The results start with an analysis that focuses on a control run where all parameters are set to their mean values. The primary goal of the control run is to assess the role of oceanic conditions on larval dispersal. Then, we present the results of the uncertainty propagation in two analysis. The first is a uni-dimensional exploration of the biological parameters space where only one variable changes at a time. The second analysis is a multi-dimensional investigation that quantifies the impacts of the combined biological uncertainties on dispersal with the goal of determining the relative importance of each input parameter for the model output uncertainties. Discussion and conclusions are drawn in the following section. The results of this study identify and rank the uncertain input parameters that contribute the most to the uncertainty in the connectivity estimates; this ranking can be used to prioritize the list of future experimental studies needed to improving connectivity estimates.

Material and methods

Biophysical model of dispersal

The CMS (Paris *et al.*, 2013b) is a coupled biophysical model used to simulate the dispersal of *A. saxatilis* larvae in the Florida Keys. This type of model uses the outputs of a physical oceanographic model, here the HYbrid COordinate Model (HYCOM; Chassignet *et al.*, 2009), to simulate the trajectories of biological particles moving under the influence of various physical and biological mechanisms. Such connectivity model is a useful tool to study the potential dispersal of planktonic organisms and the CMS has been used to estimate the connectivity of marine populations in several works including Holstein *et al.* (2014), Kitchens *et al.* (2017), Truelove *et al.* (2017), and Kough *et al.* (2019). A major feature of the CMS is its ability to simulate biological traits and behaviours so that their impacts on larval dispersal and populations connectivity can be investigated (Staaterman *et al.*, 2012; Faillettaz *et al.*, 2017). The CMS is composed of three interconnected modules: a physical oceanographic module, a habitat module, and a behavioural/biological module. The biological module of the publicly available CMS was enhanced to model larvae orientation. The details of these enhancements will be summarized in the section Biological and behavioural models.

South Florida configuration

The oceanic conditions are provided by the Florida Straits, South Florida and the Florida Keys Hybrid Coordinate Ocean Model (FKEYS-HYCOM), an oceanographic model developed to resolve the mesoscale (and also aspects of the submesoscale) circulation around South Florida and the Florida Keys island chain (Kourafalou and Kang, 2012). The model grid extends from 22.18°N to 27.51°N and from 78.0°W to 84.5°W with a horizontal resolution of 1/100° (∼900 m). The model data are available with a temporal resolution of 6 h for a continuous period from 2004 to the present. The study period of our experiment runs from May to October

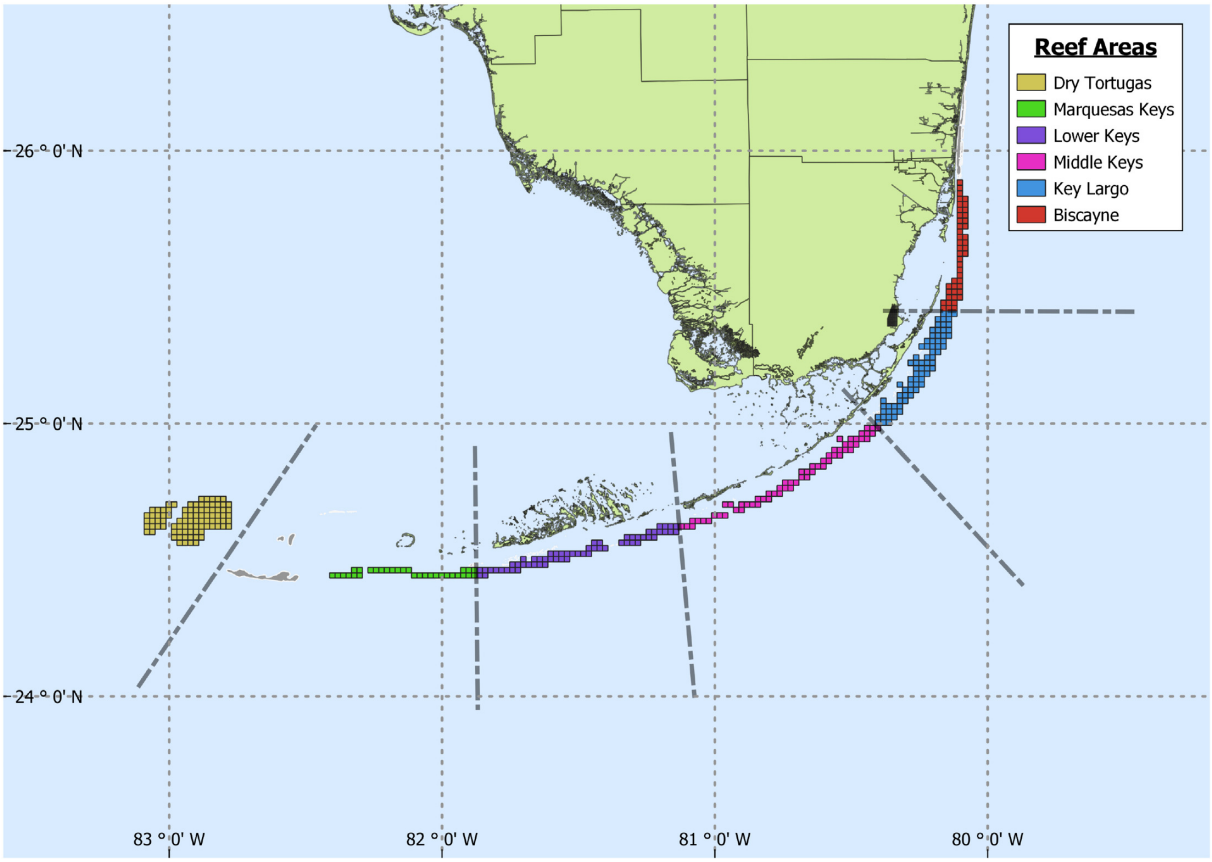

Figure 1. Map of the Florida Reef Tract. The reef habitat used in the biophysical simulations is composed of 360 grid cells each 4 km^2 in area. The reef polygons overlay the geolocalized reef map produced by the Coral Millenium Mapping Project (Andréfouët *et al.*, 2006). Reef polygons are grouped in six reef areas to facilitate the interpretation of the model outputs: DT, the MK, the LK, the MidK, KL, and B (respectively from Southwest to Northeast, marked with different colours, see box insert).

2007 (spawning period for *A. saxatilis*). While general larval supply is known to depend strongly on oceanic conditions (Banks *et al.*, 2007; Snyder *et al.*, 2014), our experiment focuses only on the biological uncertainties, and therefore, requires fixed oceanic environmental conditions. The CMS is run offline, and the motion of the particles is integrated using a 4th order Runge–Kutta method with a time step of 1200 s.

The CMS is used to simulate the dispersal of *A. saxatilis* larvae over the span of the Florida Reef Tract, USA. The coral reefs are represented by 360 grid cells each 4 km^2 in area. The reef map is subdivided into six reef areas depicted in Figure 1: Dry Tortugas (DT), the Marquesas Keys (MK), the Lower keys (LK), the Middle Keys (MidK), Key Largo (KL), and Biscayne (B). The grid cells, or reef polygons, are built by overlaying the geolocalized reef map produced by the Coral Reef Millenium Mapping Project (Andréfouët *et al.*, 2006), and supplemented by information from surveys conducted by the Southeast Fisheries Science Center, NOAA (Ault *et al.*, 2005). Coral reefs from DT to B, with reported presence of *A. saxatilis* between 2004 and 2008, are used to localize spawning and settlement grounds for this species. From May to September 2007, 100 larvae are released daily from the center of each grid cell at a depth of 10 m, in accordance with observations of reproductive behaviour (McAlary and McFarland, 1993; Aishuth *et al.*, 1998). No uncertainty is associated with the release depth to limit the complexity of the analysis.

The biological module is configured to represent *A. saxatilis*, the sergeant major, a common species of damselfish from the Pomacentridae family. This species is found in loose aggregations on shallow coral reefs in the Subtropical Atlantic and all along the Florida Reef Tract (Fishelson, 1970). This species is selected because multiple biological and behavioural traits of Pomacentridae larvae have been quantified by observations and experiments: swimming abilities (Fisher *et al.*, 2000; Bellwood and Fisher, 2001; Hogan and Mora, 2005), ontogenetic development (Robertson, 1988), pelagic larval duration (PLD; Aishuth *et al.*, 1998), hatching behaviours (Chaput *et al.*, 2019b), and ontogenetic vertical migrations (Irisson *et al.*, 2010; Corell *et al.*, 2012). Furthermore, damselfish species have been widely used as a model to study larval dispersal and settlement (Irisson *et al.*, 2004; Leis *et al.*, 2007; Staaterman *et al.*, 2012; Snyder *et al.*, 2014; Vaz *et al.*, 2016; Berenshtein *et al.*, 2018). Biological and behavioural traits used to model the larvae of *A. saxatilis* are presented in the next section.

Biological and behavioural models

Ontogenetic vertical migration

The ontogenetic vertical migrations of *A. saxatilis* are represented by a matrix of probability of vertical distribution for a given developmental age (Staaterman *et al.*, 2012). Larvae start their pelagic journey at the depth of release and migrate downward following the increase of their abilities during de-

Table 1. Ontogenetic vertical migration matrix used to represent the vertical distribution of *A. saxatilis* larvae during dispersal. The fraction of larvae present at each depth is dependent on the age of the larvae. For each age class, the sum over all the depths is equal to one.

Depth *m*	Hatching Day 0	Pre-flexion Days 1–9	Flexion Days 10–16	Post-flexion Days 16–34
5	0.10	0.05	0.05	0.01
10	0.80	0.55	0.35	0.19
30	0.10	0.30	0.32	0.33
50	0.00	0.06	0.16	0.25
70	0.00	0.03	0.09	0.18
90	0.00	0.01	0.03	0.04
$\sum$	1	1	1	1

velopment (Table 1). Depth and age distributions are based on field observations and modelling studies of dispersal of Pomacentridae larvae (Cha *et al.*, 1989; Hendriks *et al.*, 2001; Huebert *et al.*, 2010; Staaterman *et al.*, 2012).

Orientation behaviours

The CMS is enhanced with an orientation module inspired by Codling *et al.* (2004) and Staaterman *et al.* (2012). The present section describes this module, especially the uncertain parameters that modulate orientation. The aim of this orientation behaviour is to simulate fish larvae actively swimming towards the reefs at the end of their dispersal to find a settlement habitat (Atema *et al.*, 2002; Simpson *et al.*, 2004; Paris *et al.*, 2013a). The orientation model assumes that the movement of the larvae depends on (i) their swimming speed, (ii) their orientation accuracy, (iii) the position of the nearest reef, and (iv) their ability to detect the distant reefs.

When larvae are not orienting—early in their life and/or far from the reefs—their displacements are estimated by summing the horizontal velocity of the ocean currents $\vec{u}_{\text{cur}}$ and the stochastic velocity $\vec{u}_{\text{turb}}$ due to unresolved turbulence. Note that the horizontal diffusivity is set to 0.7 $m^2\,s^{-1}$ while the vertical diffusivity is set to 0.005 $m^2\,s^{-1}$, in accordance with the diffusion diagrams from Okubo (1971). The larva velocity $\vec{u}$ is then defined as

$$\vec{u} = \vec{u}_{\text{cur}} + \vec{u}_{\text{turb}}. \tag{1}$$

When larvae are orienting they will tend to swim towards the reefs, therefore, their displacements depend on the ocean currents velocity as well as their orientation velocity $\vec{u}_{\text{orient}}$. Nevertheless, the turbulence effects are assumed to be negligible in comparison to the orientation velocity (Staaterman *et al.*, 2012). In brief, the larva velocity $\vec{u}$ is then defined as

$$\vec{u} = \vec{u}_{\text{cur}} + \vec{u}_{\text{orient}}, \text{ if } age \geq \textbf{flexion} \text{ and } d \leq \beta, \tag{2}$$

where the orientation ability is controlled by two parameters: the **flexion** parameter (in days after hatch), which controls when larvae start to orient, and the $\boldsymbol{\beta}$ parameter (in km), which is the detection distance threshold. The orientation behaviour is activated when the larva sensory system and swimming abilities are sufficiently developed (controlled by *age*), around the flexion stage, and when d, the distance to the nearest reef, is smaller than the detection distance.

The orientation velocity is assumed to be horizontal and is modelled by using the larvae swimming speed Sw (cm s^{-1}) and the orientation heading θ,

$$\vec{u}_{\text{orient}} = Sw \times \begin{pmatrix} \cos\theta \\ \sin\theta \end{pmatrix}. \tag{3}$$

As per observations (Bellwood and Fisher, 2001), the larvae swimming speed increases with age and reaches a limit as described by the near-linear relationship proposed by Fisher and Bellwood (2003) and previously used by Staaterman *et al.* (2012),

$$Sw(age) = \boldsymbol{Sw}_{\text{hatch}} + (\boldsymbol{Sw}_{\text{settle}} - \boldsymbol{Sw}_{\text{hatch}})^{\frac{log_{10} age}{log_{10} \boldsymbol{PLD}}}, \tag{4}$$

where $\boldsymbol{Sw}_{\text{hatch}}$ denotes the hatching swimming speed (cm s^{-1}), $\boldsymbol{Sw}_{\text{settle}}$ the settlement swimming speed (cm s^{-1}), and ***PLD*** the pelagic larval duration (days after hatch).

At each time step, the orientation heading θ (rad) of the larva is randomly picked from a von Mises distribution (which is the circular analogue of the normal distribution),

$$f(\theta) = \frac{\exp\left(\boldsymbol{\kappa}\cos(\theta - \theta' - \mu_\zeta)\right)}{2\pi I_0(\boldsymbol{\kappa})}, \tag{5}$$

where $\boldsymbol{\kappa}$ is the orientation accuracy (dimensionless), θ' the heading at the previous time (rad), μ_ζ the mean turning angle (rad), and I_0 the modified Bessel function (of the first kind) of order 0. The orientation accuracy controls the standard deviation of the distribution $f(\theta)$ and its impact on the larvae dispersal can be appreciated from Figure 2 where the spread of trajectories is plotted for $\boldsymbol{\kappa} = 0.5$ (wider) and $\boldsymbol{\kappa} = 5$ (tighter). The mean turning angle of the larvae is modulated by the strength of the orientation cues and expressed as

$$\mu_\zeta = -\left(1 - \frac{C}{\boldsymbol{\beta}}\right) \times (\theta' - \theta_{\text{reef}}), \tag{6}$$

where C (km) is the distance from the nearest reef, $\boldsymbol{\beta}$ (km) the maximum detection distance, and θ_{reef} the heading towards the nearest reef. Headings are computed with the Haversine formula using the latitudes and longitudes of the nearest reef and the positions of the larva at time t and $t-1$, as schematically illustrated in Figure 3.

Larvae mortality and settlement

The mortality rate in the CMS is expressed as an exponential decay (Paris *et al.*, 2007; Houde, 2008):

$$A_{t+1} = A_t \times \exp\left(-\frac{2ln(2)}{\boldsymbol{PLD}} \cdot t\right), \tag{7}$$

where A_{t+1} is the proportion of larvae alive at time $t+1$.

Larvae can settle in the reefs polygons if they are competent, a period starting as early as 20 days after hatching, marked by the **competency** parameter and continuing until the end of the PLD. Larvae must be within the boundaries of a reef polygon during the competency period to settle. If larvae reach the end of the PLD without settling, or are being advected outside of the domain, they die and are removed from the system. The PLD parameter controls, therefore, both mortality during dispersal and maximum competency period.

Input parameters and their distributions

A total of seven biological and behavioural input parameters of the model (in bold in Equations (4)–(7)) are considered uncertain in this study, namely the hatching swimming speed, the settlement swimming speed, the orientation accuracy, the flexion (defined as the starting age of orientation), the maximum

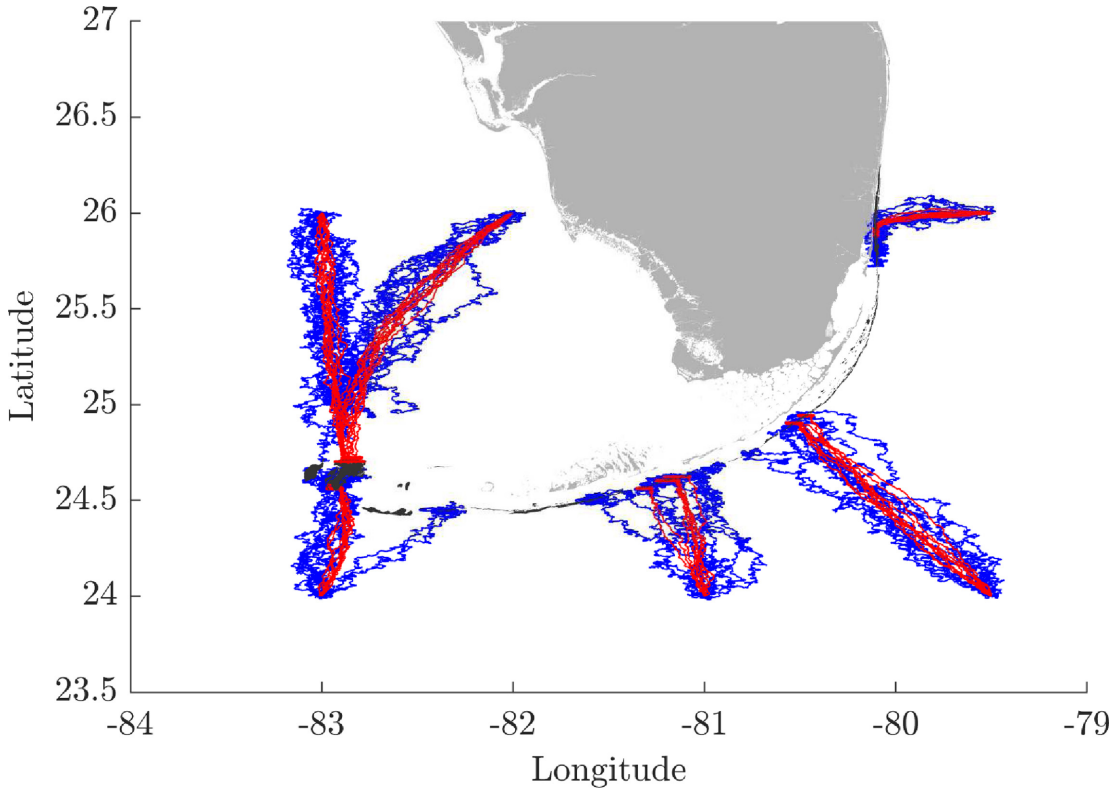

Figure 2. Simulated dispersal (10 larvae per release point) in absence of currents to illustrate the effect of the orientation accuracy κ on the spread of the trajectories. The trajectories in blue and red correspond to $\kappa = 0.5$ and $\kappa = 5$, respectively. The detection distance β is set to 5000 km.

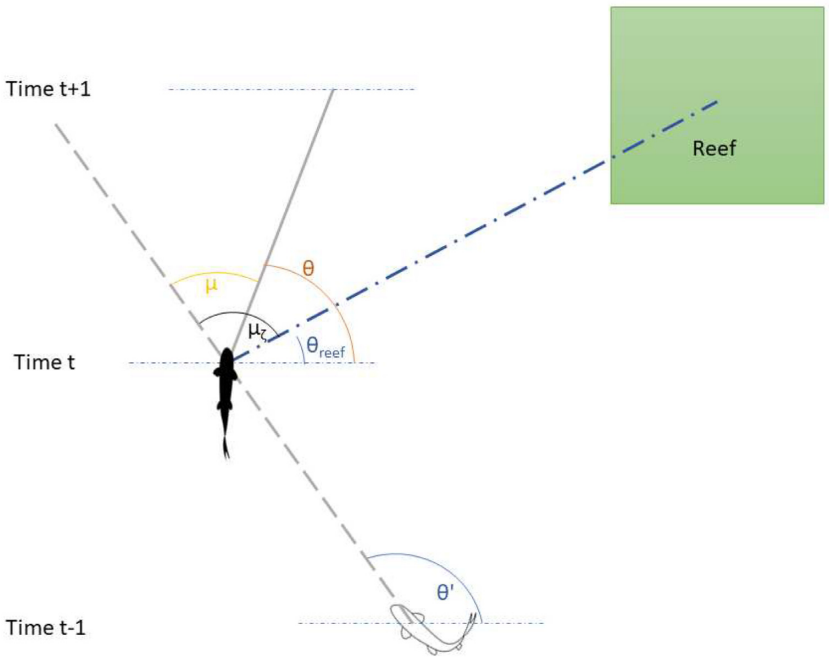

Figure 3. Computation of the heading of the larvae with the orientation behaviour. The heading θ at time t depends on the heading θ' at the previous time step $t-1$ and the direction of the reef θ_{reef}. The angle μ is the actual turning angle of the larva approximating $\mu_\zeta = \theta' - \theta_{\text{reef}}$, the optimal turning angle. In this illustration, the reef and trajectories are not represented to scale.

reef detection distance, the competency (defined as the starting age of possible settlement), and the PLD. The ranges of the uncertain input parameters (summarized in Table 2) are estimated from a literature review on *A. saxatilis*. The values are derived from experimentation on fish larvae for the swimming speeds, the flexion age, the competency age, and the PLD (Aishuth *et al.*, 1998; Bellwood and Fisher, 2001; Hogan *et al.*, 2007) and from theoretical modelling studies for the de-

Table 2. List of the uncertain input parameters in the modelling of the dispersal of *A. saxatilis*. These parameters are assumed to be independent and uniformly distributed. The ensemble of simulations relies on the ranges while the control run uses the average values of the parameters. The flexion, competency, and PLD are in days after hatch. Range values for the swimming speeds reflect the mean plus or minus standard deviation and not minimum and maximum reported values as for the other parameters.

Inputs	Function	Control run	Ranges	References
Sw_{hatch}	Swimming speed [Equation (4)]	2.395 cm s^{-1}	[2.02, 2.77] cm s^{-1}	Bellwood and Fisher (2001)
Sw_{settle}	Swimming speed [Equation (4)]	30.86 cm s^{-1}	[17.54, 44.18] cm s^{-1}	Hogan *et al.* (2007)
κ	Orientation accuracy [Equation (5)]	2.75	[0.5, 5]	Codling *et al.* (2004)
Flexion	Start orientation	13.5 d	[10, 17] d	Aishuth *et al.* (1998)
β	Detection distance [Equation (6)]	5 km	[0, 10] km	Staaterman *et al.* (2012)
Competency	Start settlement	22 d	[20, 24] d	Robertson (1988)
PLD	PLD [Equation (7)]	30 d	[28, 32] d	Aishuth *et al.* (1998)

tection distance and the orientation accuracy (Codling *et al.*, 2008; Staaterman *et al.*, 2012). We consider the maximum and minimum observed or theorized values for the parameters. The swimming speed is the only exception, we use the reported mean plus or minus one standard deviation instead of extreme values given the tendencies of flume experiments to overestimate the swimming speed sustained in the natural environment by fish larvae (Leis and Carson-Ewart, 1997). In the first part of this study, for the control run of the CMS, averaged values of the input parameters are used. The second part of this study aims at investigate the parameters space in order to quantify the impact of the uncertainties in the model.

Quantities of interest

The dispersal—the sum of larval hatching, transport, survival, and settlement (Pineda *et al.*, 2007)—estimated by the CMS is analyzed using two metrics: the connectivity matrix and a set of dispersal kernels.

The connectivity matrix represents the probability of connection between reefs using a matrix $\mathcal{M}$ of size $n \times n$ (where n is the number of reefs in the system). The source reefs are present on the y-axis while the settlement reefs are present on the x-axis. Each element $\mathcal{M}_{ij}$ indicates the fraction of larvae born in j and that settle in i. In particular, the diagonal line ($j = i$) indicates local retention (Burgess *et al.*, 2014). Exchanges between reefs are presented using the proportion of settlement (settlers over total release) on a logarithmic scale to highlight differences between areas.

The dispersal kernel of a given reef measures the probability that a larva born on that reef settles at x km from it. A dispersal kernel can be analyzed in a variety of ways depending on the focus of the study. Here, we distinguish three intervals: from 0 to 10 km (range of local retention), from 10 to 100 km (maximum distance of interest for ecological studies), and higher than 100 km (distance of interest for evolutionary studies). The analysis of the connectivity is presented here for two reefs, DT and LK, selected because they present two distinct dispersal kernels; while the results for two other reefs with significant recruitment levels, MK and MidK, are made available in the Appendix. The two remaining reefs, KL and B, are not presented in this study due to their similarities with the results obtained for the LK and MidK albeit overall shorter dispersal distances and lower recruitment levels.

Impact of oceanic circulation on dispersal

The oceanic features responsible for successful dispersal are highlighted by a combined analysis of the time-series of settlement and the hyperbolic Lagrangian Coherent Structures (LCS). The LCS fields are used to identify the boundaries of coherent structures, such as eddies, in the flow. LCS are useful to predict and understand the transport of biological particles in the oceans (Olascoaga *et al.*, 2008; Kai *et al.*, 2009; Haller, 2015). Here, the LCS are diagnosed with the Finite Time Lyapunov Exponent (FTLE), a measure of the exponential rate of separation of trajectories of particles over a finite time interval (van Sebille *et al.*, 2018). The trajectories of particles dispersing in the system tend to follow attracting LCS that act as transport barriers (van Sebille *et al.*, 2018). We obtain the trajectories by uniformly releasing particles over the Eulerian velocity fields of the FKEYS-HYCOM hydrodynamic model and backtracking (or foretracking) their trajectories over 14 d. Attracting LCS correspond to the ridges of the FTLE field calculated using backward integration (repulsing LCS, not shown in the current study, are calculated using forward integration). LCS are analyzed in correlation with settlement rates to identify the mechanisms responsible for high and low settlement.

Uncertainty propagation

Principle

The biological and behavioural parameters used in the model (see Table 2 for the ranges) are uncertain because of diverse reasons including simplifications and idealizations, difficulties in directly measuring these constants in the field, differences between laboratory and field conditions, as well as inter-individual variations. The aim of the following section is to quantify the uncertainties in the model output, the probability of settlement, caused by uncertainties in the aforementioned parameters using a probabilistic framework. The first step in this process is the specification of the distribution of the uncertain inputs. For this study, we assume that the distribution of the input parameters over their individual ranges are uniform and independent. This choice yields equiprobable realizations over the parametric domain.

The second step consists in forward propagating the input distributions to compute the corresponding output distribution. This is accomplished in practice with an ensemble of runs that sample the parametric domain using for instance a Monte Carlo method. Each ensemble member, or realization, has a fixed value for each uncertain parameter; the CMS can then be run and its output, the probability of settlement,

calculated. The statistical analysis can proceed once enough samples have been collected so as to estimate the output distributions. A reliable analysis requires that enough samples (thousands to millions) be collected to gain confidence in the output statistics.

The ensemble run constitutes the majority of computational cost, particularly for large expensive models. The tension between mitigating the computational cost and improving the statistical analysis can be resolved by introducing a surrogate model (also referred to as an emulator or a meta-model) that approximates, everywhere in the parametric domain, the changes in the model output (connectivity matrix or dispersal kernel) caused by the changes in the model inputs. The surrogate is implemented non-intrusively by considering the direct model as a "black-box" and exploiting the input–output relationships of the output quantity at the ensemble members. The surrogate model can take several forms and here we rely on the PC approach detailed below. If the surrogate is sufficiently accurate, it can be used instead of the original model to carry on the statistical analysis. For example, directly computing the statistical moments of the model outputs, performing the variance analysis, and estimating the probability density function of the model outputs.

PC description

Let $\boldsymbol{\xi}$ be the random vector collecting the seven uncertain input parameters. A PC surrogate $f_p(\boldsymbol{\xi})$ of a model output $f(\boldsymbol{\xi})$ (Le Maitre and Knio, 2010; Iskandarani *et al.*, 2016) is essentially a polynomial expansion of the form:

$$f(\boldsymbol{\xi}) \approx f_p(\boldsymbol{\xi}) = \sum_{k=0}^{p} \widehat{f}_k \psi_k(\boldsymbol{\xi}), \tag{8}$$

where $p+1$ denotes the number of terms in the truncated series. The model output considered in our study is the probability of settlement at a given distance, the metric used to compute the dispersal kernels. The PC expansion provides a functional representation of the output in which $\widehat{f}_k$ are the deterministic series coefficients and $\psi_k(\boldsymbol{\xi})$ are orthogonal (multivariate) polynomial basis functions in $\boldsymbol{\xi}$. In practice, the components of the vector $\boldsymbol{\xi}$ are linearly scaled from their physical ranges listed in Table 2 to the dimensionless interval $[-1, 1]$. The scaling for the detection distance β is for example $\xi_5 = 2(\beta - \beta_{\min})/(\beta_{\max} - \beta_{\min}) - 1$. Because all the uncertain input parameters are uniformly distributed, the orthogonal basis functions are the products of univariate Legendre polynomials (A more detailed explanation of the orthogonal polynomial basis associated to each input distribution is presented in Iskandarani *et al.* (2016) and Sudret (2014)).

The PC coefficients $\widehat{f}_k$ are determined with the ordinary least squares method to filter out the noise produced by the intrinsic stochasticity in the CMS, stemming from the turbulence effect and the biased correlated random walk used in the behavioural model. The PC coefficients are then determined by minimizing the sum of the squared residuals at the N predetermined training points $\boldsymbol{\xi}_i$, namely

$$\| f - f_p \|_{\mathrm{LS}}^2 = \sum_{i=1}^{N} \left(f(\boldsymbol{\xi}_i) - f_p(\boldsymbol{\xi}_i) \right)^2 . \tag{9}$$

The least squares solution for the PC coefficients $\widehat{\boldsymbol{f}}$ is

$$\widehat{\boldsymbol{f}} = \left(\boldsymbol{\Psi}\boldsymbol{\Psi}^{\top} \right)^{-1} \boldsymbol{\Psi} \boldsymbol{f}, \tag{10}$$

Table 3. Number of parameters of the PC surrogates used in the unidimensional (1D) and the multidimensional (5D) analysis. The 1D analysis explores the space of the input parameters individually while the 5D analysis explores the space of all the input parameters and their interactions. The training points are used to compute the coefficients of the surrogates and the validation points are selected independently to quantify the fit of the models.

	1D	5D
Number of PC coefficients	5	95
Maximum PC degree	4	7
Number of training points	5	351
Number of validation points	4	100

where the design matrix $\boldsymbol{\Psi} = [\psi_k(\boldsymbol{\xi}_i)]$ has a size $(p+1) \times N$, and the vector of outputs $\boldsymbol{f} = [f(\boldsymbol{\xi}_i)]$ has a size N (the number of training points).

When using the least squares approach, random samplings (e.g. Monte Carlo, quasi-Monte Carlo, or Latin Hypercube Sampling) are commonly chosen to estimate the PC coefficients. The truncation strategy then relies on the total order of the polynomial n_{ord}, leading to $p + 1 = (n_{dim} + n_{ord})!/(n_{dim}!n_{ord}!)$ PC coefficients, where n_{dim} is the number of uncertain parameters. However, we rely here on sparse grid points that provide a well-conditioned matrix $\boldsymbol{\Psi}\boldsymbol{\Psi}^{\top}$. To be brief, a sparse grid is based on a quadrature rule associated with partial tensorization and is classically associated to a set of PC that are exactly integrated when using the sparse pseudo-spectral projection method (Constantine *et al.*, 2012; Conrad and Marzouk, 2013). Traditional full tensorization Gauss quadratures maximize the accuracy of the integration per sample but are not suitable when multiple dimensions are considered because of the large number of sampling points needed. In our example, with five dimensions and a total polynomial order 4, we would need $(n_{ord}+1)^{n_{dim}} = 3125$ sample points (One run taking approximately 1 h of parallel computation on 45 CPU cores of Intel Sandy Bridge E5-2670 (2.6 GHz)). Instead, we use a (level 3) sparse grid that contains 351 points and allows us to compute the 95 coefficients of the truncated PC expansions.

Simulation scenarios

Working on the base provided by the control run, we conducted several ensembles of simulations to quantify the impact of the uncertain input parameters. These ensembles of simulations are organized as follows:

1. Unidimensional analysis, also called one-at-a-time analysis , to investigate the impact of varying the 7 input parameters one by one and to quantify their impact on the connectivity. As a result of this unidimensional analysis, we decided to remove two parameters, hatching swimming speed and PLD, for the rest of the experiment. A validation test is performed for each unidimensional study, and each test consisted of four independent validation points (Table 3).
2. Multidimensional analysis, also called global sensitivity analysis, to investigate the combined impact of five biological input parameters. The contribution of each uncertain input parameter, along with the contribution of their interactions, to the variance in settlement are further quantified with the Sobol indices (Sobol, 1993). The

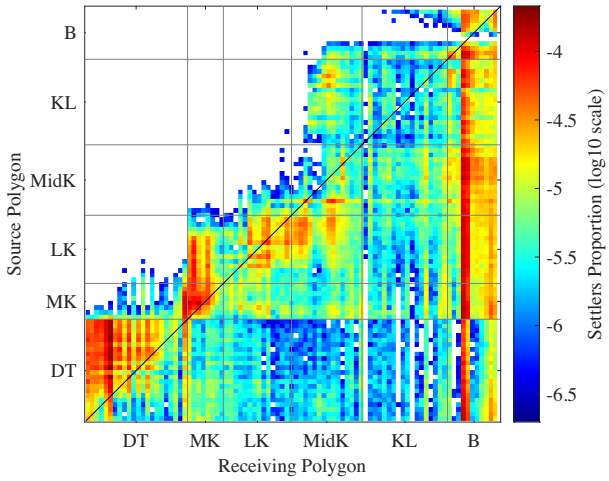

Figure 4. Control run probability connectivity matrix in the Florida Keys presenting the proportion of larvae transitioning from sources/spawning sites (*y*-axis) to settlement sites (*x*-axis). The reefs polygons are grouped in six geographical areas: DT, MK, LK, MidK, KL, and B. Local retention within a population is represented on the diagonal, while upstream and downstream connectivity are represented above and below the diagonal line, respectively. The proportion of settlement, in logarithmic scale, is averaged over four neighboring polygons for downscaling and smoothing the connectivity matrix (from 360 to 90 pixels).

accuracy of the multi-dimensional surrogate is assessed using 100 validation points (Table 3) randomly selected using latin hypercube sampling (near-random sampling in multi-dimensional case). The surrogate validation can be found in the Appendix.

Results of the control run

Spatial analysis of settlement

The control run presents the estimated larval connectivity associated with the average behavioural and biological traits, whereby the biological parameters listed in Table 2 are set to their mean values. The resulting connectivity matrix in Figure 4 shows the probability of exchanges between all the reef-polygons in the Florida Reef Tract system. We observe a dominant downstream dispersal (below the diagonal) in accordance with the Southwest–Northeast direction of the Florida Current within the western part of the Straits of Florida. Larvae spawned in the Southwestern part of the Keys (DT, MK, and LK) end up settling in the reefs located further North. A total of three reefs have important local retention (around the diagonal line): DT, MK, and LK. The northern reef B is a sink, receiving more larvae than sending away. This feature is explained by its location at the northern end of the reef system dominated by a downstream transport, which accumulates competent larvae over B in our model. The MidK and KL reefs are characterized by larger export than import of larvae. The hot spots of settlement in our estimate of connectivity are located in DT, MK, and B. Finally, there is a non-negligible upstream flux of larvae, especially from KL which delivers larvae to the MidK, and the LK that sends settlers to the MK.

In addition to the connectivity matrix, we analyse the dispersal kernel that provides the individual characteristic of a reef. The dispersal kernel from DT (Figure 5a) is characterized by two separate peaks with a high probability of dispersal, especially for short distances, reflecting local retention since no other reef is located within 50 km. Local retention and exports to other reefs are separated on the dispersal kernel by a break in settlement probability due to the absence of suitable habitat between DT and the MK. Further away from the source reefs, the proportion of settlement fluctuates slightly around 1×10^{-4} between 50 and 300 km. Successful dispersal probability increases to 5×10^{-4} after 300 km and is limited to 330 km (larvae reaching the northern most reefs on B). The LK dispersal kernel (Figure 5b) is characterized by a bi-modal distribution, with a first peak of successful dispersal over short to medium distances (between 0 and 75 km), and a second over large dispersal distances (between 175 and 225 km). The highest proportion of settlement (5.7×10^{-4}) is located at 25 km and the highest proportion at long dispersal distance (4.5×10^{-4}) is found at 180 km. Probabilities of successful dispersal between 75 and 175 km are characterized by lower values around 1.3×10^{-4}. The maximum distance of dispersal from the LK is 240 km. The dispersal kernels for MK and the MidK can be found in appendix (Figure A1a and b).

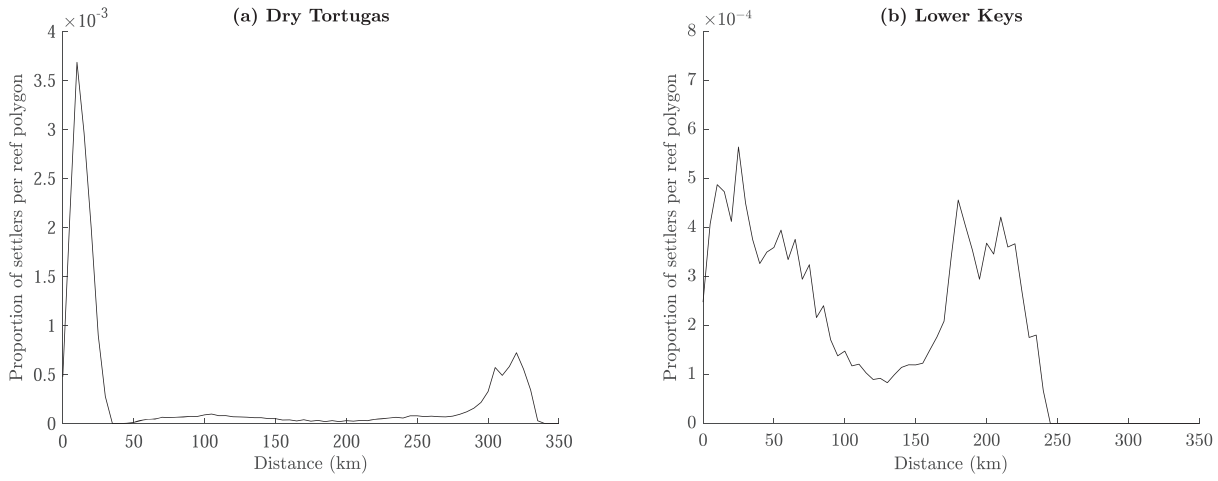

Figure 5. Control run dispersal kernels from (a) DT and (b) LK reefs.

Temporal analysis of settlement

Analysis of the temporal pattern of release of successful—future settlers—larvae (red bars on Figure 6) and settlement (blue bars on Figure 6) shows multiple distinct peaks of release followed by similar peaks of settlement spread over the reefs. The time-lag between release and settlement is about 22 d, which corresponds to the beginning of the competency period, meaning that larvae typically settle shortly after becoming competent in the model. Because the daily release of larvae is constant, fluctuations in settlement are due to the influence of the currents during larvae dispersal. Larvae released over DT settle either back to DT, or seem to be advected by the Florida Current and settle on the reefs downstream over B (Figure 6a–h). Larvae released over the LK seem to be advected either upstream in June and settle over the MK and back to the LK, or downstream and settle over the other reefs (Figure 6i–n). There is almost no connection with DT, except for one settlement event at the beginning of June.

Lagrangian Coherent Structures

The spatial representation of the LCS, coupled with an analysis of the settlement time-series, shows which configuration of the currents enhances successful larvae dispersal. Figure 7(a)–(d) show four snapshots of FTLE fields showing the features responsible for high and low settlement rates. The ridges of the backward-time FTLEs correspond to attracting LCS positions during the selected 14 d time-span.

The first FTLE field (Figure 7a) highlights the presence of an eddy north of DT, just preceding the peak of settlement in DT observed at the end of June (Figure 6). Such mesoscale eddies (with characteristic length scales larger than 30 km; Kourafalou and Kang, 2012) are known to act as a mechanism to entrain larvae and retain them, as they are advected along the Florida Current, eventually delivering them to downstream reefs (Sponaugle *et al.*, 2005; Vaz *et al.*, 2016). This eddy traps larvae released from DT and drive them towards the Florida Keys reefs, where they are released due to eddy breaking through interaction with the island topography, or through splitting to smaller eddies (Kourafalou and Kang, 2012). When this eddy is not present (Figure 7b), larvae are advected further away from the reefs, which results in a low settlement event. We can also observe another eddy south of DT in Figure 7(a). This eddy travels with the Florida Current from Southwest to Northeast. It retains larvae and can distribute them all along the reefs of the Florida Keys when its path nears the reef tract (the eddy trajectory is controlled by the meandering of the Florida Currents; Kourafalou and Kang, 2012).

Figure 7(c) shows two eddies travelling along the coast at the end of July. These eddies are associated with high settlement rates for larvae released over the LK at the end of July and the beginning of August (Figure 6). Larvae originating from the Southern Keys (Marquesas and LK) and from DT are carried by these eddies and retained over the reefs of the upper Keys. The dynamic of this type of eddy activity may explain the upstream connectivity observed in the connectivity matrix (Figure 4). When, however, these eddies travel further away offshore as the Florida Current meanders away from the Florida Keys (Figure 7d), they retain larvae far from the reefs, resulting in low recruitment periods. These eddies often converge towards the reefs around B, before moving North along the Gulf Stream. This dynamic could explain why the reefs in B act as sink in our simulations.

Results of uncertainty analysis

Unidimensional analysis

Analysis of the impact of individual input parameters on the output estimates of connectivity are used to classify the relative importance of each parameter for the different reefs. As mentioned before, three settlement distances are considered in the analysis: from 0 to 10 km, from 10 to 100 km, and higher than 100 km (Figure 8). These dispersal distances have different ecological implications and consequences for local fish populations, and need to be analyzed separately.

Of all the six reef areas, DT is the one where input parameters induce the most important variance in the estimates of settlement (see y-scale of left panels of Figure 8). For this reef, we observe that the impact of the input parameters is mostly restricted to the short dispersal distances (blue curves). The main influential parameter on the proportion of settlement is the detection distance of the reefs (β), followed by the orientation accuracy (κ). Unsurprisingly, both are positively correlated with the proportion of settlement. In contrast, there are negative correlations between the proportion of settlement and the flexion (beginning of orientation) and competency (beginning of settlement) parameters. The probability of dispersal over short distance dispersal decreases when the larvae are forced to remain in the pelagic realm for longer, delaying settlement. Finally, the hatching swimming speed (Sw_{hatch}), the settlement swimming speed (Sw_{settle}), and the PLD do not seem to influence the settlement of larvae released in DT in our simulations when varied individually in the range of values tested.

The right panels of Figure 8 show that the LK exhibit a different response to the analysis. We observe for this reef that the proportion of settlement varies similarly over short and long distances, but settlement is one order of magnitude smaller than DT (see the y-axis scale). Larvae released over the reefs of the LK have higher probability of settling over short and long dispersal distances (Figure 5) and the impact of input parameters is more important on these dispersal ranges (Figure 8, right panel). The proportion of settlement is positively correlated with the detection distance and the orientation accuracy (as for DT), but this time it is also positively correlated with the settlement swimming speed. In contrast, we observe a negative correlation for the competency and flexion parameters. Again, variations of the hatching swimming speed and the PLD have a limited impact on the proportion of settlement in the context of this model.

The results of the unidimensional analysis on the two other reefs (Figure A2) are comparable to the LK results but with a higher variance level of the proportion of settlement. As for DT and LK, the hatching swimming speed and the PLD have a weak influence on the results. Analysis shows that these two input parameters produce the smallest output variance for three out of the four reefs considered (Table A1). Therefore, these parameters are held fixed to their mean values for the following multi-dimensional analysis. It is likely that the interaction of these two parameters with the five others would increase their impact on the output variance. However, this decision is motivated by the need to decrease the number of samples necessary to assess all of the interaction between parameters.

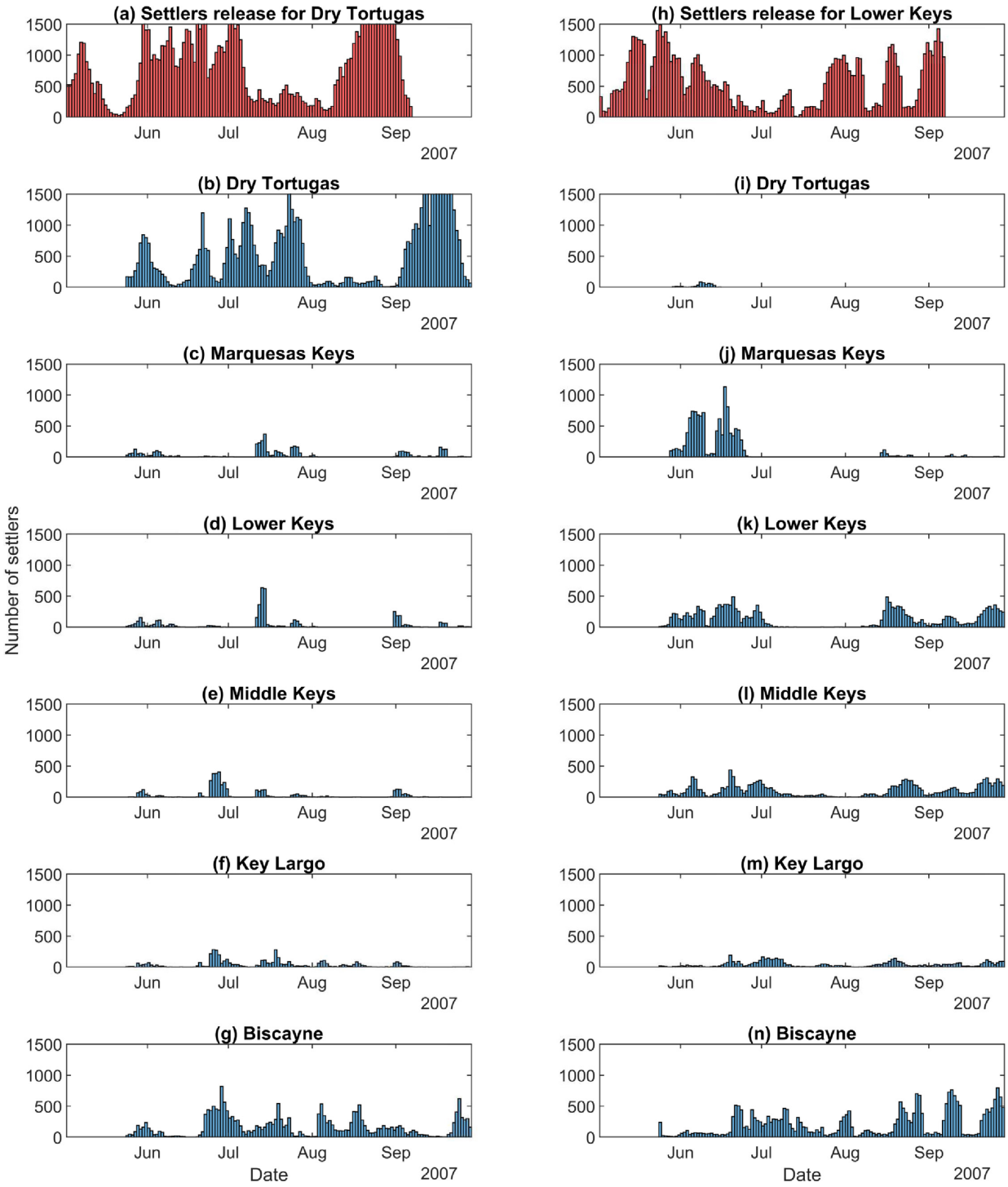

Figure 6. (a) Time series of settlers release (red) over DT, and (b)–(g) time-series of settlement over the Florida Keys Reefs (blue) between May and September 2007. (h) Time series of settlers release (red) over the LK, and (i)–(n) time-series of settlement for these larvae (blue). Peaks of release of successful larvae are followed by peaks of settlement with a time lag due to the pre-competency period.

Multidimensional analysis

This section presents the uncertainty analysis of the combined effects of the five (remaining) uncertain input parameters on the dispersal kernels. Figure 9 shows the dispersal kernels mean with plus or minus the standard deviation estimated by the PC surrogates for DT and the LK after accounting for the uncertainties in the five uncertain input parameters. The shape of the dispersal kernels, produced by the PC surrogates, remains similar to the estimates produced by the control run (also plotted), meaning that variation in input parameters do not affect qualitatively the distribution of the dispersal kernels in our model. However, the mean dispersal kernel shows that the control run overestimates the proportion of settlers especially for long distances where the control dispersal kernel is

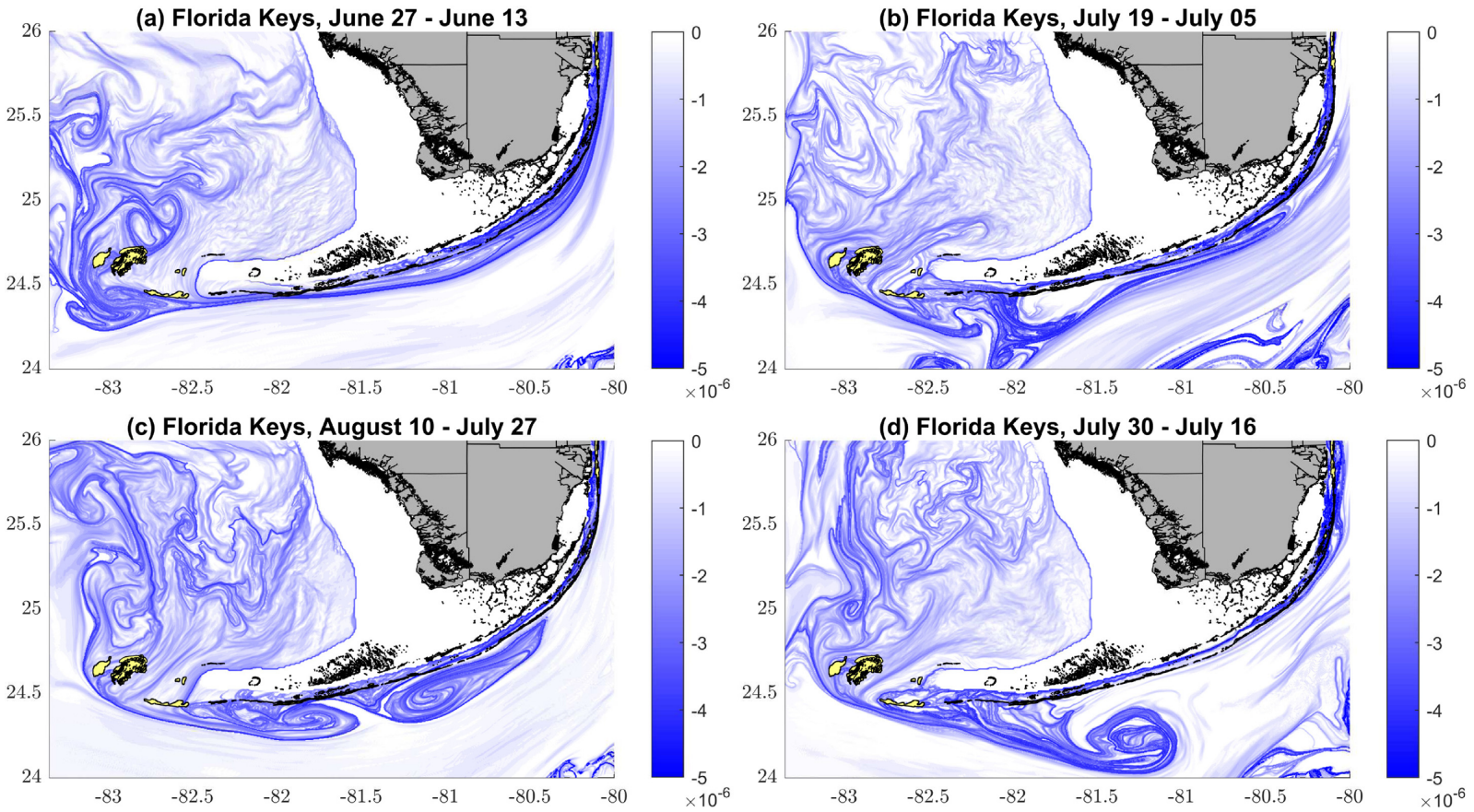

Figure 7. LCS in the Florida Keys in 2007. Backward-time FTLEs are computed over 14 d at a depth of 30 m for date corresponding to a high (a) and low (b) recruitment events for the reefs of DT, and for a high (c) and low (d) recruitment events for the reefs of the LK. Note the presence of a mesoscale eddy approaching DT from the North (a), the presence of two mesoscale eddies on top of the reefs of the Marquesas and the LK (c), and one eddy moving offshore, away from the reefs (d).

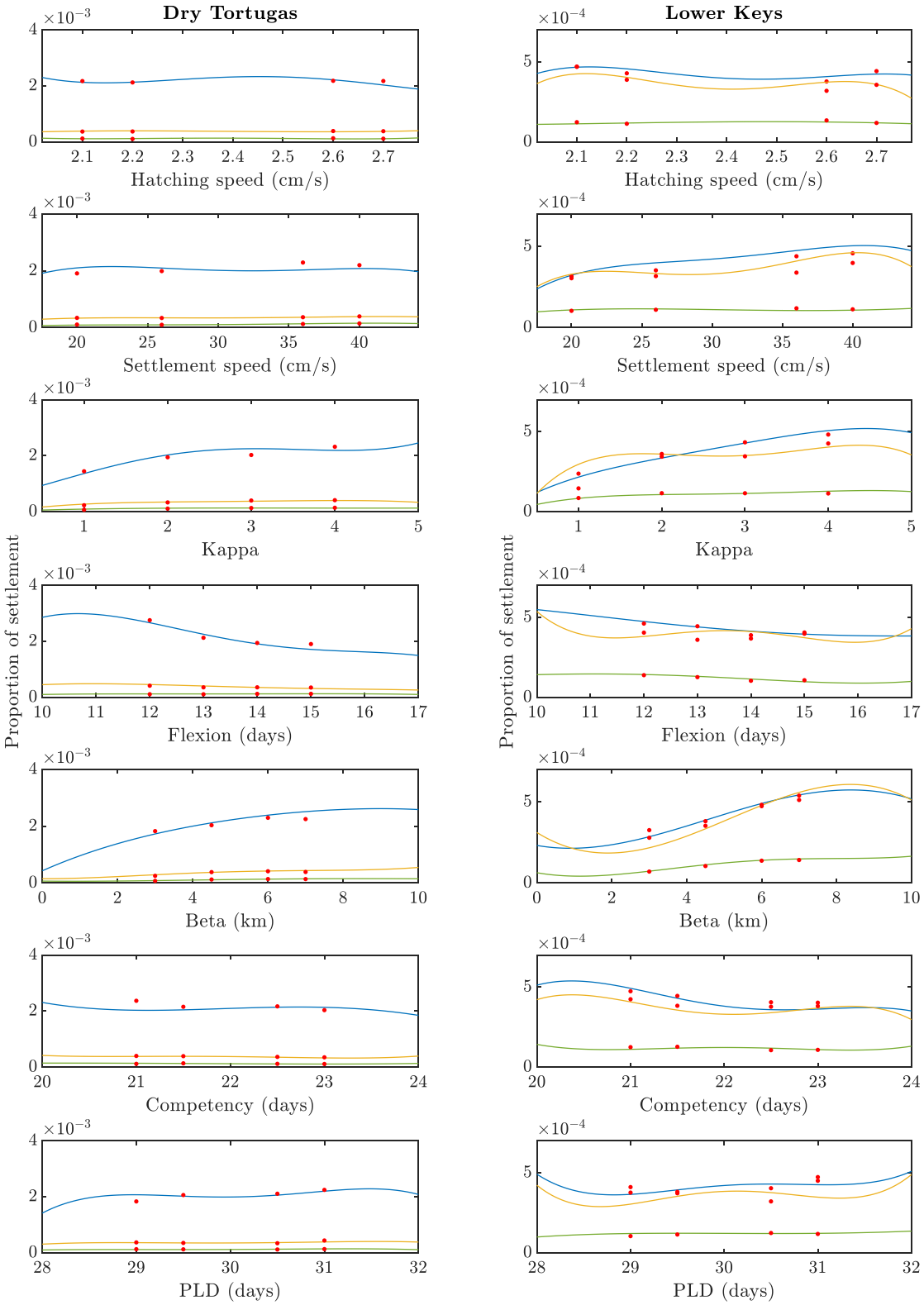

Figure 8. Variation of the proportion of settlement as predicted by the PC surrogate within the seven input parameter ranges for three dispersal distances: short (0–10 km: blue), medium (10–100 km: orange), and long (>100 km: green). Results are presented for larvae released from DT (left) and the LK (right). Input parameters are varied one-at-a-time, with the values of the other non-varied parameters fixed to their mean. Red points represent the settlement as predicted by 28 validation runs of the CMS and show that the surrogate model gives a close approximation of the CMS predictions.

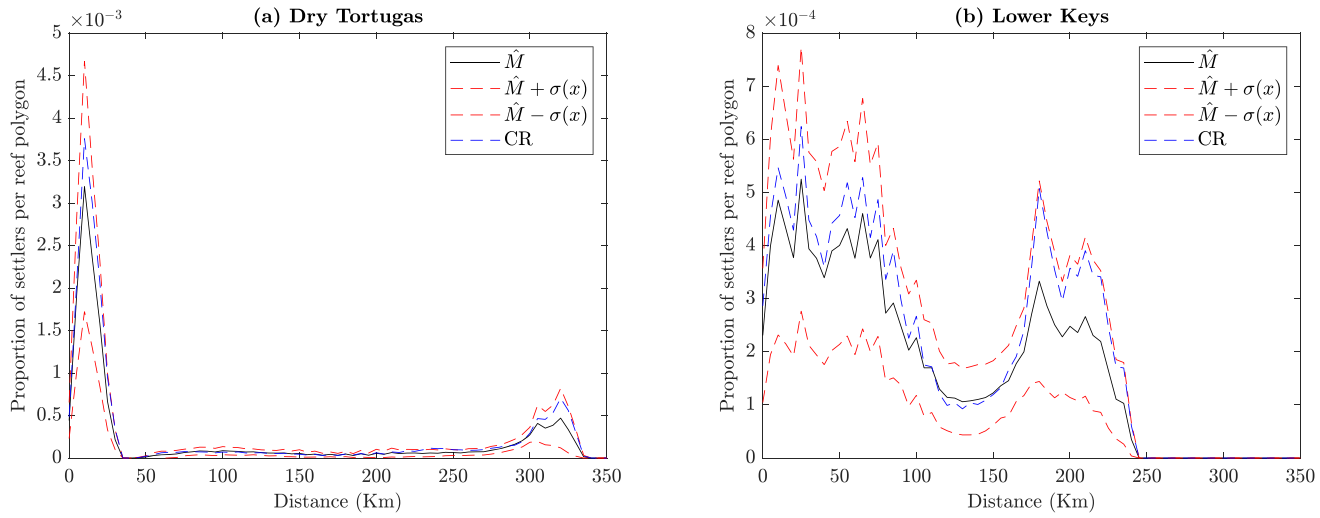

Figure 9. PC statistical moments of dispersal kernels from (a) DT and (b) LK reefs: mean dispersal $\hat{M}$ and standard deviation σ. Dispersal kernels means have similar shapes as the control run dispersal kernels for both reefs. Average proportion of settlement is lower than control run estimate (CR: blue dotted line) across most dispersal distances, highlighting an overestimation in the control scenario. Standard deviations for both reefs are larger for short and long distance dispersal, matching the peaks of the distributions, and show larger impact of larval behaviours in these regions.

roughly superimposed on the dispersal kernel mean plus one standard deviation. The standard deviation is roughly proportional to the mean; we observe large values for the high peaks of settlement meaning that parameter uncertainties mainly impact the dispersal distances with high settlement. Similar conclusions can be drawn for the LK (Figure 9b), and both MK and MidK (Figure A3a and b).

Analysis of variance

A global sensitivity analysis of the variance is performed to identify the dominant contributors to the dispersal kernel variance. Variation in the standard deviation of the dispersal kernels shows that the impact of the input parameters depends on the distance of dispersal. The analysis of variance is, therefore, performed by considering the three dispersal distance groups established earlier (short (0–10 km), medium (10–100 km), and long (> 100 km)) to estimate the first order and total sensitivity indices. The so-called Sobol indices decompose the variance of the model outputs into fraction attributed to the input parameters.

The first-order sensitivity index measures the variance in the model output caused by the uncertainty in each input parameters acting on its own. In our model, the first-order sensitivity index shows that the detection distance (β) is the most influential parameter in the proportion of settlement for all the dispersal distances and from both reefs of interest (Figure 10a and c). Interestingly, β has an opposite behaviour for DT and the LK. For DT, the impact of β on the variance decreases with the distance of dispersal (explaining from 17% down to 5% of the total variance). In contrast, for the LK, the impact of β on the variance increases with the dispersal distances (explaining from 12% up to 29% of the total variance). This means that better orientation behaviour retains more larvae over DT, but favours the export of larvae from the LK, an effect which could be due to the different oceanographic regimes over these two reefs (see the subsection on LCS). The other input parameters explain less than 10% of the total variance. For both reefs, only a small fraction of the total variance of settlement is explained by the first-order sensitivity index.

The total-order index measures the contribution to the output variance of an input parameter, including all variance caused by its interactions, up to the fourth-order here, with the other input parameters. Analysis of Figure 10(b) and (d) highlights the importance of interactions between parameters. Indeed, a larger fraction of the settlement variance is explained by the total-order index than the first-order index. The input parameter β—and its interactions with the other parameters—remains the main driver of variability for both reefs, explaining from 40 to 29% of the variance for DT (decreasing with distance), and from 35 to 47% for the LK (increasing with distance). At short dispersal distances from DT, the Flexion parameter—and its interactions—explains 28% of the variance while the orientation accuracy (κ) explains 19% of it. For the two other dispersal distances, the remaining variance is explained equally (around 20%) by the four input parameters, excluding β. For dispersal from the LK, κ—and its interactions—is the parameter explaining the most variance after β. Its importance decreases with the dispersal distance. At long dispersal distance (> 100 km) the parameter Flexion explains as much variance as κ (around 18%). The input parameters Swimming speed and Competency period explain less than 10% of the total variance.

The two other reefs exhibit a similar trend as the LK (Figure A4). More variance is explained by interactions between input parameters than by the first-order sensitivity indices. The input parameter β explains a large fraction of the variance for all three dispersal distances, while the remaining variance is shared between the other four parameters with a higher contribution of the parameters κ and Flexion.

Discussion

Biophysical models are subject to many uncertainties that hinder our ability to accurately simulate the dispersal of fish larvae. While uncertainties are unavoidable, they must be identified and quantified (Ådlandsvik *et al.*, 2009). We demonstrate how this can be done using the PC approach. First, uncertainties in input parameters are characterized through their

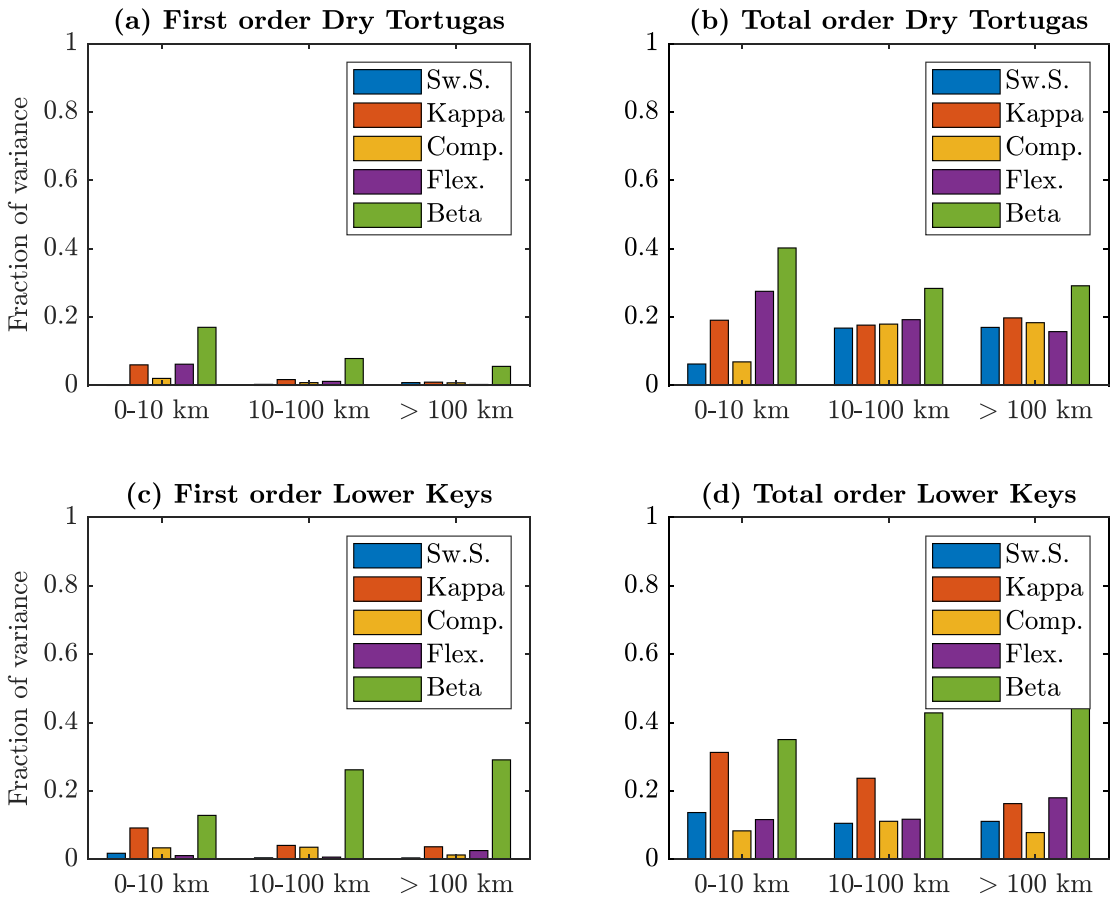

Figure 10. (a) and (c) Sobol indices of first order ($S_{1...5}$) and (b) and (d) total order sensitivity ($T_{1...5}$) for the input parameters: Settlement swimming speed, κ, Competency, Flexion, and β. Variances induced by the parameters are presented for DT (a) and (b) and the LK (c) and (d) at three dispersal ranges: short (0–10 km), medium (10–100 km), and long (> 100 km).

ranges using the available data on *A. saxatilis*; second these uncertainties are propagated through the connectivity model to construct a surrogate model that is subsequently analyzed to extract the uncertainties in the model forecasts. We generally find that uncertainties in the dispersal kernels are due to a few understudied behavioural traits and that the impacts of biological uncertainties depend on the release location of the larvae.

The dispersal kernels for *A. saxatilis* in the Florida Keys depend on larvae behaviours and on larvae reactions to oceanic conditions. The supply of *A. saxatilis* larvae is conditioned on the presence of two types of oceanic features, which control the location of attracting LCS over the reefs: an eddy located in the vicinity of DT and the propagation of mesoscale eddies along the Florida Reef Tract (Kourafalou *et al.*, 2009; Kourafalou and Kang, 2012). These eddies aggregate and retain the larvae over the reef during their dispersal, and therefore increase settlement. In our model, as well as in previous modelling studies of fish and coral larvae dispersal, pulses of settlement are observed in conjunction with the displacement of attracting LCS over the reefs (Vaz *et al.*, 2016; Frys *et al.*, 2020). Observed arrival of Pomacentridae larvae on the reefs in the Florida Keys is also correlated with the propagation of these eddies (Sponaugle *et al.*, 2005, 2012; D'Alessandro *et al.*, 2007). Interestingly, both Sponaugle *et al.* (2005) and D'Alessandro *et al.* (2007) observed a similar number of eddies for three different years as in our study (approximately three mesoscale eddies associated with peaks in larval settlement), confirming that the year selected for our study is representative of the oceanic circulation in the Florida Keys. The settlement predicted by our simulations also matched the arrival of *Stegastes partitus* (another common Pomacentridae) larvae observed in 2007 by Sponaugle *et al.* (2012). Larger settlement rates were observed over the LK than the Upper Keys (equivalent to KL in our study). However, our simulations do not show the relatively large settlement observed in the LK in the first half of August 2007 (Sponaugle *et al.*, 2012).

The analysis of the probabilistic dispersal kernels shows that the maximum detection distance threshold, the orientation accuracy, and the development of the orientation behaviour are the parameters, together with their interaction, responsible for most of the variation in settlement of *A. saxatilis*. It is likely that the dispersal of species with similar orientation behaviours would also be sensitive to these uncertainties. The orientation behaviour parameters are poorly constrained by previous observations and their values remain hypothetical. The maximum detection distance is thought to vary depending on the type of orientation cues. Here, we assumed that olfactory and/or acoustic cues are coming from the reef. Studies have shown that both are used by coral reef fish larvae during dispersal and settlement (Lecchini *et al.*, 2005; Gerlach *et al.*, 2007; Simpson *et al.*, 2011; Gordon *et al.*, 2018). While odors from the reefs could only spread to hundreds of meters (Atema, 2006; Paris *et al.*, 2013a), sounds could reach

much further, on the order of tens of kilometers (Staaterman, 2015). However, the distance at which these cues are detected remains poorly known. The uncertainties associated with this parameter can be decreased with further experiments on fish larvae orientation. More experiments would also be required to decrease the uncertainties associated with the beginning of orientation. Pomacentridae larvae have well-developed olfactory and auditory abilities at the pre-settlement stages (Wright *et al.*, 2005), and studies on other species have found functional sensory systems as soon as hatching occurs (Hu *et al.*, 2019). Larvae could potentially orient early during dispersal, which could increase retention rate and settlement (Staaterman *et al.*, 2012); therefore, the development of orientation behaviour need to be studied more in depth. Finally, the orientation accuracy is responsible for some variations in the dispersal kernel estimates. The uncertainties in this parameter are compounded by our limited knowledge of its actual distribution range in fish larvae. Experimental work on the orientation of fish larvae have shown the wide disparity of behaviours between tested individuals (see for examples Cresci *et al.* (2019b), Berenshtein *et al.* (2014), and Leis *et al.* (2014)). However, these studies have traditionally focused on the mean direction of orientation rather than the spread of directions, which would be more informative to estimate the larvae's orientation accuracy.

Interestingly, the uncertainties in swimming speed and PLD did not affect the dispersal kernels substantially in the context of our model. Development of the swimming abilities has been extensively described for various coral reef fish species (Bellwood and Fisher, 2001; Clark *et al.*, 2005; Hogan *et al.*, 2007). The uncertainties associated with this parameter in our model are largely caused by inter-individual variability. Previous modelling studies have shown that the mere presence of swimming ability increases settlement rates and, overall, influences the connectivity (Leis *et al.*, 2007; Drake *et al.*, 2018; Faillettaz *et al.*, 2018; Cresci *et al.*, 2020b). Changes in swimming speeds (both at hatching and settlement), however, do not significantly influence the dispersal kernels in the context of our study. This could either mean that inter-individual differences in swimming speed do not influence connectivity in this region with strong currents, or that the model chosen to represent the larval orientation decreases the importance of the swimming speed. Indeed, the swimming speed only acts on short distances—determined by the detection distance—at the end of dispersal. Over these short distances, differences introduced by the variation in swimming speed could be too small to make a difference. On the other hand, these small variations in swimming speed could be important for other functions during dispersal. Slightly higher swimming speed could be beneficial during feeding activity, to avoid predators, or to maintain group cohesion (Mackenzie and Kiørboe, 1995; Chaput *et al.*, 2019a), which would influence the larval supply. The limited impact of the PLD on settlement in our model could be linked to the interaction between mortality and orientation behaviour. Because of the mortality, very few larvae remain at the end of the competency period. In addition, we find that most larvae settle as soon as they enter their competency period. Therefore, variance in PLD has a limited impact on the estimated settlement. The PLD, however, has been found to have large effect on settlement in models without orientation behaviours (Treml *et al.*, 2012; Andrello *et al.*, 2013), and the limited impact could be specific to our case study on *A. saxatilis*.

The PC estimates show that the variability in biological and behavioural traits does not affect the shape of the dispersal kernels, but rather the relative number of settlers per dispersal distance. We hypothesize that the shape of dispersal kernels is mainly due to the interaction with the oceanic currents, the distribution of suitable habitat, and finally the type of orientation behaviours that larvae use during dispersal. Here, we tested only one type of orientation behaviour with the biased correlated random walk model. Our goal was to represent a direct orientation mechanism that would be based on the detection of olfactory or acoustic cues coming from the reefs. This type of orientation has been documented, or hypothesized, for many coral reef fish larvae (Lecchini *et al.*, 2005; Wright *et al.*, 2005; Paris *et al.*, 2013a; Hu *et al.*, 2019). However, it is not the only type of orientation used by larvae. Indeed, previous studies have documented the use of magnetic compass (Bottesch *et al.*, 2016; Cresci *et al.*, 2019b) or celestial orientation (Mouritsen *et al.*, 2013; Faillettaz *et al.*, 2015; Cresci *et al.*, 2019a) by fish larvae. As larvae react diversely to different environmental cues during dispersal, the orientation mechanism could impact significantly the connectivity patterns and modify the dispersal kernels. Furthermore, larvae could also use more than one type of orientation cue at different stages of dispersal, depending on the environmental context (Cresci *et al.*, 2020a). In this case, more complex behavioural models would be needed to accurately simulate larval dispersal. However, it is necessary to improve our understanding of the environmental cues followed by the larvae during dispersal before increasing the complexity of biophysical models, especially since our study shows that the highest uncertainties are associated with orientation behaviour.

The probabilistic approach adopted here increases the reliability of the model using a twofold advantage. First, it quantifies the impact of the combined uncertain inputs on the output, providing the user with an estimate of the model usefulness. Second, the PC approach reduces the computational burden through its surrogate and simplifies the subsequent statistical analysis by avoiding a large Monte Carlo ensemble. Both methods, Monte Carlo and PC, rely on sampling to compute the output statistics. However, whereas Monte Carlo treats each sample independently, PC methods attempt to find a relationship between samples. In both cases, the computational cost rises linearly with the number of realizations N. However, the error for Monte Carlo sampling decreases by $\sqrt{1/N}$ for the mean, whereas it decreases exponentially fast ($1/N^N$) for PC methods, meaning that fewer samples are required for a similar accuracy. The number of realizations needed to achieve a certain accuracy for PC, however, depends on the degree of the polynomial used and the number of uncertain parameters (the so-called curse of dimensionality): the number of samples can increase exponentially fast with the number of dimensions. Fortunately, more sophisticated sampling strategies, like the sparse quadrature used in this study, can reduce this number tremendously.

Other methods have been used to quantify the uncertainties in biophysical models. Linear regression models synthesize the impact of a large number of input variables (Treml *et al.*, 2012; Torrado *et al.*, 2021), but are limited to linear relationships and lack the precision offered by the PC surrogate estimates. Multi-factor, multi-variable analysis of variance (MANOVA) can similarly be used to quantify the combined impact of parameters but requires large number of simulations (Edwards *et al.*, 2007). The advantage of the PC procedure, over those

methods, is that it produces a surrogate that is essentially a polynomial series that lends itself to straightforward analysis and simplifies the estimation of the mean, variance, and Sobol indices (Alexanderian *et al.*, 2012). Finally, the PC approach is limited to models with continuous inputs and outputs distributions, and therefore, could not be used to compare the impact of different orientation behaviours (Fox *et al.*, 2016; Fobert *et al.*, 2019), or to estimate ecosystem-wide connectivity where multiple dispersal strategies cohabit (Treml *et al.*, 2015; Gary *et al.*, 2020). Regardless of the method chosen, similar sensitivity studies should be conducted to understand how dispersal is modulated by the orientation behaviours used by fish larvae of different species, and in different areas of the world. Complete validation, or rejection, of biophysical models might not be attainable (Oreskes *et al.*, 1994), but uncertainty quantification in modelled and empirical estimates of connectivity (Kaplan *et al.*, 2017; D'Aloia *et al.*, 2018) will increase the confidence in our ability to correctly describe larval dispersal.

Authors' contributions

RC designed the study, implemented and ran the model, analyzed the data, interpreted the results, and wrote the manuscript. PS analyzed the data, interpreted the results, and wrote the manuscript. PM designed the LCS study, analyzed the data, and wrote the manuscript. VHK provided the oceanographic data, interpreted the results, and wrote the manuscript. MI designed the study, implemented the model, interpreted the results, wrote the manuscript, and funded the research.

Conflict of interest

The authors declare no competing interests.

Data availability

The data underlying this article will be shared on reasonable request to the corresponding author. Source codes for the Connectivity Modelling System (CMS) are available at https://github.com/beatrixparis/connectivity-modeling-system. Codes for the orientation module are available at https://github.com/RomainChaput/connectivity-modeling-system. Finally, codes for the uncertainty quantification, and a sample of the model outputs are available at https://github.com/RomainChaput/Uncertainty-estimation-in-biophysical-model-of-connectivity.

Acknowledgements

This research comprises a portion of RC's doctoral thesis requirements (University of Miami). The simulations were performed with resources provided by the University of Miami Center for Computational Science. We thank HeeSook Kang for help with the oceanographic data and Natalie Perlin for help with the CMS. We thank Maria Josefina Olascoaga and David M. Kaplan for their constructive discussions on the subject of LCS and biological uncertainties. Finally, we thank an anonymous reviewer and Marco Andrello for constructive comments and suggestions, which greatly improved our paper.

References

Ådlandsvik, B., Bartsch, J., Brickman, D., Browman, H. I., Edwards, K., Fiksen, Ø., Gallego, A *et al.* 2009. Manual of recommended practices for modelling physical – biological interactions during fish early life. ICES Cooperative Research Report, 295: 1–111.

Aishuth, S. R., Tucker, J. W. J., and Hateley, J. 1998. Egg and larval development of laboratory-reared sergeant major, *Abudefduf saxatilis* (Pisces, Pomacentridae). Bulletin of Marine Science, 62: 121–133.

Alexanderian, A., Winokur, J., Sraj, I., Srinivasan, A., Iskandarani, M., Thacker, W. C., and Knio, O. M. 2012. Global sensitivity analysis in an ocean general circulation model: a sparse spectral projection approach. Computational Geosciences, 16: 757–778.

Almany, G. R., Planes, S., Thorrold, S. R., Berumen, M. L., Bode, M., Saenz-Agudelo, P., Bonin, M. C. *et al.* 2017. Larval fish dispersal in a coral-reef seascape. Nature Ecology and Evolution, 1: 1–7.

Andréfouët, S., Muller-Karger, F. E., Robinson, J. A., Kranenburg Christine, J., Torres-Pulliza, D., Spraggins, S. A., and Murch, B. 2006. Global assessment of modern coral reef extent and diversity for regional science and management applications : a view from space. *In* Proceedings of the 10th International Coral Reefs Symposium, pp. 1732–1745.

Andrello, M., Jacobi, M. N., Manel, S., Thuiller, W., and Mouillot, D. 2015. Extending networks of protected areas to optimize connectivity and population growth rate. Ecography, 38: 273–282.

Andrello, M., Mouillot, D., Beuvier, J., Albouy, C., Thuiller, W., and Manel, S. 2013. Low connectivity between Mediterranean marine protected areas: a biophysical modeling approach for the dusky grouper *Epinephelus marginatus*. PLoS ONE, 8: e68564.

Atema, J. 2006. Chemical signals in the marine environment: dispersal, detection, and temporal signal analysis. Proceedings of the National Academy of Sciences of the United States of America, 92: 62–66.

Atema, J., Kingsford, M. J., and Gerlach, G. 2002. Larval reef fish could use odour for detection, retention and orientation to reefs. Marine Ecology Progress Series, 241: 151–160.

Ault, J. S., Bohnsack, J. A., Smith, S. G., and Luo, J. 2005. Towards sustainable multispecies fisheries in the Florida, USA, coral reef ecosystem. Buttetin of Marine Science, 76: 28.

Ayata, S. D., Ellien, C., Dumas, F., Dubois, S., and Thiébaut, É. 2009. Modelling larval dispersal and settlement of the reef-building polychaete *Sabellaria alveolata*: role of hydroclimatic processes on the sustainability of biogenic reefs. Continental Shelf Research, 29: 1605–1623.

Banks, S. C., Piggott, M. P., Williamson, J. E., Bové, U., Holbrook, N. J., and Beheregaray, L. B. 2007. Oceanic variability and coastal topography shape genetic structure in a long-dispersing sea urchin. Ecology, 88: 3055–3064.

Bellwood, D. R., and Fisher, R. 2001. Relative swimming speeds in reef fish larvae. Marine Ecology Progress Series, 211: 299–303.

Berenshtein, I., Kiflawi, M., Shashar, N., Wieler, U., Agiv, H., and Paris, C. B. 2014. Polarized light sensitivity and orientation in coral reef fish post-larvae. PLoS ONE, 9: 1–9.

Berenshtein, I., Paris, C. B., Gildor, H., Fredj, E., Amitai, Y., and Kiflawi, M. 2018. Biophysical simulations support schooling behavior of fish larvae throughout ontogeny. Frontiers in Marine Science, 5: 1–12.

Bottesch, M., Gerlach, G., Halbach, M., Bally, A., Kingsford, M. J., and Mouritsen, H. 2016. A magnetic compass that might help coral reef fish larvae return to their natal reef. Current Biology, 26: R1266–R1267.

Burgess, S., Nickols, K., Griesemer, C., Barnett, L., Dedrick, A., Satterthwaite, E., Yamane, L *et al.* 2014. Beyond connectivity : how empirical methods can quantify population persistence to improve marine protected area design: supplementary Information. Ecological Society of America, 24: 8.

Cha, S. S., Mcgowan, M. F., and Richards, W. J. 1989. Vertical distribution of fish larvae off the Florida Keys, 26 May–5 June 1989. Bulletin of Marine Science, 54: 828–842.

Chaput, R., Majoris, J. E., Buston, P. M., and Paris, C. B. 2019a. Hydrodynamic and biological constraints on group cohesion in plankton. Journal of Theoretical Biology, 482: 1–9.

Chaput, R., Majoris, J. E., Guigand, C. M., Huse, M., and D'Alessandro, E. K. 2019b. Environmental conditions and paternal care determine hatching synchronicity of coral reef fish larvae. Marine Biology, 166: 118.

Chassignet, E. P., Hurlburt, H. E., Metzger, E. J., Smedstad, O. M., Cummings, J. A., Halliwell, G. R., Bleck, R *et al*.. 2009. US GODAE global ocean prediction with the HYbrid Coordinate Ocean Model (HYCOM). Oceanography, 22: 49–59.

Clancy, D., Tanner, J. E., McWilliam, S., and Spencer, M. 2010. Quantifying parameter uncertainty in a coral reef model using Metropolis-Coupled Markov Chain Monte Carlo. Ecological Modelling, 221: 1337–1347.

Clark, D. L., Leis, J. M., Hay, A. C., and Trnski, T. 2005. Swimming ontogeny of larvae of four temperate marine fishes. Marine Ecology Progress Series, 292: 287–300.

Claro, R., Lindeman, K. C., Kough, A. S., and Paris, C. B. 2018. Biophysical connectivity of snapper spawning aggregations and marine protected area management alternatives in Cuba. Fisheries Oceanography, 28:1–10.

Codling, E. A., Hill, N. A., Pitchford, J. W., and Simpson, S. D. 2004. Random walk models for the movement and recruitment of reef fish larvae. Marine Ecology Progress Series, 279: 215–224.

Codling, E. A., Plank, M. J., and Benhamou, S. 2008. Random walk models in biology. Journal of the Royal Society Interface, 5: 813–834.

Conrad, P. R. and Marzouk, Y. M. 2013. Adaptive Smolyak pseudospectral approximations. SIAM Journal on Scientific Computing, 35:A2643–A2670.

Constantine, P. G., Eldred, M. S., Phipps, E. T., and Laboratories, S. N. 2012. Sparse pseudospectral approximation method. Computer Methods in Applied Mechanics and Engineering, 229–232: 1–12.

Corell, H., Moksnes, P. O., Engqvist, A., Döös, K., and Jonsson, P. R. 2012. Depth distribution of larvae critically affects their dispersal and the efficiency of marine protected areas. Marine Ecology Progress Series, 467: 29–46.

Cowen, R. K., and Sponaugle, S. 2009. Larval dispersal and marine population connectivity. Annual Review of Marine Science, 1: 443–466.

Cresci, A. 2020. A comprehensive hypothesis on the migration of European glass eels (*Anguilla anguilla*). Biological Reviews, 95: 1–14.

Cresci, A., Allan, B. J. M., Shema, S. D., Berit, A., and Browman, H. I. 2020a. Orientation behavior and swimming speed of Atlantic herring larvae (*Clupea harengus*) in situ and in laboratory exposures to rotated artificial magnetic fields. Journal of Experimental Marine Biology and Ecology, 526: 151358.

Cresci, A., Durif, C. M., Paris, C. B., Thompson, C. R. S., Shema, S., Skiftesvik, A. B., and Browman, H. I. 2019a. The relationship between the moon cycle and the orientation of glass eels (*Anguilla anguilla*) at sea. Royal Society Open Science, 6:6190812.

Cresci, A., Paris, C. B., Foretich, M. A., Durif, C. M., Shema, S. D., O'Brien, C. E., Vikebø, F. B *et al*. 2019b. Atlantic haddock (*Melanogrammus aeglefinus*) larvae have a magnetic compass that guides their orientation. iScience, 19: 1173–1178.

Cresci, A., Sandvik, A. D., Sævik, P. N., Ådlandsvik, B., Olascoaga, M. J., Miron, P., Durif, C. M *et al*.. 2020b. The lunar compass of European glass eels (*Anguilla anguilla*) increases the probability that they recruit to North Sea coasts. Fisheries Oceanography, 00: 1–16.

D'Alessandro, E., Sponaugle, S., and Lee, T. 2007. Patterns and processes of larval fish supply to the coral reefs of the upper Florida Keys. Marine Ecology Progress Series, 331: 85–100.

D'Aloia, C. C., Xuereb, A., Fortin, M., Bogdanowicz, S. M., and Buston, P. M. 2018. Limited dispersal explains the spatial distribution of siblings in a reef fish population. Marine Ecology Progress Series, 607: 143–154.

Drake, P. T., Edwards, C. A., Morgan, S. G., and Satterthwaite, E. V. 2018. Shoreward swimming boosts modeled nearshore larval supply and pelagic connectivity in a coastal upwelling region. Journal of Marine Systems, 187: 96–110.

Edwards, K. P., Hare, J. A., Werner, F. E., and Seim, H. 2007. Using 2-dimensional dispersal kernels to identify the dominant influences on larval dispersal on continental shelves. Marine Ecology Progress Series, 3522: 77–87.

Faillettaz, R., Blandin, A., Paris, C. B., Koubbi, P., and Irisson, J. O. 2015. Sun-compass orientation in mediterranean fish larvae. PLoS ONE, 10: 1–15.

Faillettaz, R., Durand, E., Paris, C. B., Koubbi, P., and Irisson, J.-O. 2017. Swimming speeds of Mediterranean settlement-stage fish larvae nuance Hjort's aberrant drift hypothesis. Limnology and Oceanography, 63:14–61.

Faillettaz, R., Paris, C. B., and Irisson, J.-O. 2018. Larval fish swimming behavior alters dispersal patterns from marine protected areas in the North-Western Mediterranean Sea. Frontiers in Marine Science, 5: 1–12.

Fishelson, L. 1970. Behaviour and ecology of a population of *Abudefduf saxatilis* (Pomacentridae, Teleostei) at Eilat (Red Sea). Animal Behaviour, 18: 225–237.

Fisher, R., and Bellwood, D. R. 2003. Undisturbed swimming behaviour and nocturnal activity of coral reef fish larvae. Marine Ecology Progress Series, 263: 177–188.

Fisher, R., Bellwood, D. R., and Job, S. D. 2000. Development of swimming abilities in reef fish larvae. Marine Ecology Progress Series, 202: 163–173.

Fobert, E. K., Treml, E. A., and Swearer, S. E. 2019. Dispersal and population connectivity are phenotype dependent in a marine metapopulation. Proceedings of the Royal Society B: Biological Sciences, 286: 1–9.

Foster, S. A. 1987. Diel and lunar patterns of reproduction in the Caribbean and Pacific sergeant major damselfishes *Abudefduf saxatilis* and *A. troschelii*. Marine Biology, 95: 333–343.

Fox, A. D., Henry, L. A., Corne, D. W., and Roberts, J. M. 2016. Sensitivity of marine protected area network connectivity to atmospheric variability. Royal Society Open Science, 3:311.

Frys, C., Saint-Amand, A., Le Hénaff, M., Figueiredo, J., Kuba, A., Walker, B., Lambrechts, J *et al*.. 2020. Fine-scale coral connectivity pathways in the Florida Reef tract: implications for conservation and restoration. Frontiers in Marine Science, 7: 1–16.

García-García, L. M., Ruiz-Villarreal, M., and Bernal, M. 2016. A biophysical model for simulating early life stages of sardine in the Iberian Atlantic stock. Fisheries Research, 173: 250–272.

Gary, S. F., Fox, A. D., Biastoch, A., Roberts, J. M., and Cunningham, S. A. 2020. Larval behaviour, dispersal and population connectivity in the deep sea. Scientific Reports, 10: 1–12.

Gerlach, G., Atema, J., Kingsford, M. J., Black, K. P., and Miller-Sims, V. 2007. Smelling home can prevent dispersal of reef fish larvae. Proceedings of the National Academy of Sciences of the United States of America, 104: 858–863.

Goldstein, E. D., and Sponaugle, S. 2020. Juvenile reef fish growth and survival related to subregional patterns of primary production. Marine Biology, 167: 1–10.

Gordon, T. A. C., Harding, H. R., Wong, K. E., Merchant, N. D., Meekan, M. G., McCormick, M. I., Radford, A. N *et al*.. 2018. Habitat degradation negatively affects auditory settlement behavior of coral reef fishes. Proceedings of the National Academy of Sciences of the United States of America, 115: 5193–5198.

Haller, G. 2015. Lagrangian coherent structures. Annual Review of Fluid Mechanics, 47: 137–162.

Hendriks, I. E., Wilson, D. T., and Meekan, M. G. 2001. Vertical distributions of late stage larval fishes in the nearshore waters of the San Blas Archipelago, Caribbean Panama. Coral Reefs, 20: 77–84.

Hogan, J. D., Fisher, R., and Nolan, C. 2007. Critical swimming speed of settlement stage coral reef fishes from the Caribbean: a methodological and geographical comparison. Bulletin of Marine Science, 80: 219–231.

Hogan, J. D., and Mora, C. 2005. Experimental analysis of the contribution of swimming and drifting to the displacement of reef fish larvae. Marine Biology, 147: 1213–1220.

Holstein, D. M., Paris, C. B., and Mumby, P. J. 2014. Consistency and inconsistency in multispecies population network dynamics of coral reef ecosystems. Marine Ecology Progress Series, 499: 1–18.

Houde, E. D. 2002. Mortality. *In* Fishery Science: The Unique Contributions of Early Life Stages, pp. 64–87. Ed. by L. Fuiman, and R.G. Werner, Chapter 3. Blackwell Publishing, Oxford.

Houde, E. D. 2008. Emerging from Hjort's shadow. Journal of Northwest Atlantic Fishery Science, 41: 53–70.

Hu, J., Fennel, K., Mattern, J. P., and Wilkin, J. 2012. Data assimilation with a local Ensemble Kalman Filter applied to a three-dimensional biological model of the Middle Atlantic Bight. Journal of Marine Systems, 94: 145–156.

Hu, Y., Majoris, J. E., Buston, P. M., and Webb, J. F. 2019. Potential roles of smell and taste in the orientation behaviour of coral-reef fish larvae: insights from morphology. Journal of Fish Biology, 95: 311–323.

Huebert, K. B., Sponaugle, S., and Cowen, R. K. 2010. Predicting the vertical distributions of reef fish larvae in the Straits of Florida from environmental factors. Canadian Journal of Fisheries and Aquatic Sciences, 67: 1755–1767.

Irisson, J. O., LeVan, A., De Lara, M., and Planes, S. 2004. Strategies and trajectories of coral reef fish larvae optimizing self-recruitment. Journal of Theoretical Biology, 227: 205–218.

Irisson, J. O., Paris, C. B., Guigand, C., and Planes, S. 2010. Vertical distribution and ontogenetic "migration" in coral reef fish larvae. Limnology and Oceanography, 55: 909–919.

Iskandarani, M., Wang, S., Srinivasan, A., Thacker, W. C., Winokur, J., and Knio, O. M. 2016. An overview of uncertainty quantification techniques with application to oceanic and oil-spill simulations. Journal of Geophysical Research: Oceans, 121: 2789–2808.

Kai, E. T., Rossi, V., Sudre, J., Weimerskirch, H., Lopez, C., Hernandez-Garcia, E., Marsac, F *et al.*. 2009. Top marine predators track Lagrangian coherent structures. Proceedings of the National Academy of Sciences of the United States of America, 106: 8245–8250.

Kaplan, D. M., Cuif, M., Fauvelot, C., Vigliola, L., Nguyen-Huu, T., Tiavouane, J., and Lett, C. 2017. Uncertainty in empirical estimates of marine larval connectivity. ICES Journal of Marine Science, 74: 1723–1734.

Kingsford, M. J., Leis, J. M., Shanks, A., Lindeman, K. C., Morgan, S. G., and Pineda, J. 2002. Sensory environments, larval abilities and local self-recruitment. Bulletin of Marine Science, 70: 309–340.

Kitchens, L. L., Paris, C. B., Vaz, A. C., Ditty, J. G., Cornic, M., Cowan, J. H., and Rooker, J. R. 2017. Occurrence of invasive lionfish (*Pterois volitans*) larvae in the northern Gulf of Mexico: characterization of dispersal pathways and spawning areas. Biological Invasions, 19: 1971–1979.

Kough, A. S., Belak, C. A., Paris, C. B., Lundy, A., Cronin, H., Gnanalingam, G., Hagedorn, S *et al.*. 2019. Ecological spillover from a marine protected area replenishes an over-exploited population across an island chain. Conservation Science and Practice, 1: e17.

Kough, A. S., Paris, C. B., and Butler, M. J. 2018. The spatial context of "winning" in MPA network design: location matters. Conservation Letters, 113:e12455.

Kourafalou, V. H., and Kang, H. 2012. Florida Current meandering and evolution of cyclonic eddies along the Florida Keys Reef Tract: are they interconnected? Journal of Geophysical Research, 117: 1–25.

Kourafalou, V. H., Peng, G., Kang, H., Hogan, P. J., Smedstad, O.-m., and Weisberg, R. H. 2009. Evaluation of global ocean data assimilation experiment products on South Florida nested simulations with the HYbrid Coordinate Ocean Model. Ocean Dynamics, 59: 47–66.

Le Maitre, O. and Knio, O. M. 2010. Spectral Methods for Uncertainty Quantification. Springer, Netherlands.

Lecchini, D., Shima, J., Banaigs, B., and Galzin, R. 2005. Larval sensory abilities and mechanisms of habitat selection of a coral reef fish during settlement. Oecologia, 143: 326–334.

Leis, J. M. 2006. Are larvae of demersal fishes plankton or nekton? Advances in Marine Biology, 51: 57–141.

Leis, J. M. 2007. Behaviour as input for modelling dispersal of fish larvae: behaviour, biogeography, hydrodynamics, ontogeny, physiology and phylogeny meet hydrography. Marine Ecology Progress Series, 347: 185–193.

Leis, J. M., and Carson-Ewart, B. M. 1997. In situ swimming speeds of the late pelagic larvae of some Indo-Pacific coral-reef fishes. Marine Ecology Progress Series, 159: 165–174.

Leis, J. M., Wright, K. J., and Johnson, R. N. 2007. Behaviour that influences dispersal and connectivity in the small, young larvae of a reef fish. Marine Biology, 153: 103–117.

Leis, J., Paris, C., Irisson, J., Yerman, M., and Siebeck, U. 2014. Orientation of fish larvae in situ is consistent among locations, years and methods, but varies with time of day. Marine Ecology Progress Series, 505: 193–208.

McAlary, F. A., and McFarland, W. N. 1993. The effect of light and darkness on hatching in the pomacentrid *Abudefduf saxatilis*. Environmental Biology of Fishes, 37: 237–244.

Mackenzie, B. R., and Kiørboe, T. 1995. Encounter rates and swimming behavior of pause-travel and cruise larval fish predators in calm and turbulent laboratory environments. Limnology and Oceanography, 40: 1278–1289.

Mattern, J. P., Fennel, K., and Dowd, M. 2013. Sensitivity and uncertainty analysis of model hypoxia estimates for the Texas-Louisiana shelf. Journal of Geophysical Research: Oceans, 118: 1316–1317.

Mouritsen, H., Atema, J., Kingsford, M. J., and Gerlach, G. 2013. Sun compass orientation helps coral reef fish larvae return to their natal reef. PLoS ONE, 8: 1–9.

Okubo, A. 1971. Oceanic diffusion diagrams. Deep-Sea Research and Oceanographic Abstracts, 18: 789–802.

Olascoaga, M. J., Beron-Vera, F. J., Brand, L. E., and Koçak, H. 2008. Tracing the early development of harmful algal blooms on the West Florida Shelf with the aid of Lagrangian coherent structures. Journal of Geophysical Research, 113: 1–10.

Olszewski, J., Haehnel, M., Taguchi, M., and Liao, J. C. 2012. Zebrafish larvae exhibit rheotaxis and can escape a continuous suction source using their lateral line. PLoS ONE, 75.e36661.

Oreskes, N., Shrader-Frechette, K., and Belitz, K. 1994. Verification, validation, and confirmation of numerical models in the Earth sciences. Science, 263: 641–646.

Painter, K. J., and Hillen, T. 2015. Navigating the flow: Individual and continuum models for homing in flowing environments. Journal of the Royal Society Interface, 12: 1–30.

Paris, C. B., Atema, J., Irisson, J. O., Kingsford, M., Gerlach, G., and Guigand, C. M. 2013a. Reef odor: a wake up call for navigation in reef fish larvae. PLoS ONE, 8: 1–8.

Paris, C. B., Chérubin, L. M., and Cowen, R. K. 2007. Surfing, spinning, or diving from reef to reef: effects on population connectivity. Marine Ecology Progress Series, 347: 285–300.

Paris, C. B., Cowen, R. K., Claro, R., and Lindeman, K. C. 2005. Larval transport pathways from Cuban snapper (Lutjanidae) spawning aggregations based on biophysical modeling. Marine Ecology Progress Series, 296: 93–106.

Paris, C. B., Helgers, J., van Sebille, E., and Srinivasan, A. 2013b. Connectivity modeling system: a probabilistic modeling tool for the multi-scale tracking of biotic and abiotic variability in the ocean. Environmental Modelling and Software, 42: 47–54.

Pineda, J., Hare, J., and Sponaugle, S. 2007. Larval transport and dispersal in the coastal ocean and consequences for population connectivity. Oceanography, 20: 22–39.

Prappas, J. M., Greene, L. E., and White, R. 1991. Reproductive behavior of the sergeant major, *Abudefdf saxatis*, within a closed system aquarium. Environmental Biology of Fishes, 31: 33–40 .

Robertson, D. R. 1988. Settlements and population dynamics of *Abiudeduf saxatilis* on parch reefs in Caribbean Panama. *In* Proceedings of the 6th International Coral Reef Symposium. 2: 839–844. Townsville, Australia.

Shulzitski, K., Sponaugle, S., Hauff, M., Walter, K., D'Alessandro, E. K., and Cowen, R. K. 2015. Close encounters with eddies: oceanographic features increase growth of larval reef fishes during their journey to the reef. Biology Letters, 11: 20140746.

Simpson, S. D., Meekan, M. G., Mccauley, R. D., and Jeffs, A. 2004. Attraction of settlement-stage coral reef fishes to reef noise. Marine Ecology Progress Series, 276: 263–268.

Simpson, S. D., Piercy, J. J. B., King, J., and Codling, E. A. 2013. Modelling larval dispersal and behaviour of coral reef fishes. Ecological Complexity, 16: 68–76.

Simpson, S. D., Radford, A. N., Tickle, E. J., Meekan, M. G., and Jeffs, A. G. 2011. Adaptive avoidance of reef noise. PLoS ONE, 6: 2–6.

Snyder, R. E., Paris, C. B., and Vaz, A. C. 2014. How much do marine connectivity fluctuations matter? The American Naturalist, 184: 523–30.

Sobol, I. M. 1993. Sensitivity estimates for nonlinear mathematical models. MMCE, 1: 407–414.

Sponaugle, S., Lee, T., Kourafalou, V., and Pinkard, D. 2005. Florida Current frontal eddies and the settlement of coral reef fishes. Limnology and Oceanography, 50: 1033–1048.

Sponaugle, S., Paris, C., Walter, K. D., Kourafalou, V., and D'Alessandro, E. 2012. Observed and modeled larval settlement of a reef fish to the Florida Keys. Marine Ecology Progress Series, 453: 201–212.

Staaterman, E. 2015. Temporal and spatial patterns in coral reef soundscapes and their relevance for larval fish orientation. Open Access Dissertations 1402. University of Miami Scholarly Repository, Coral Gables, FL.

Staaterman, E., and Paris, C. B. 2013. Modeling larval fish navigation: the way forward. Journal of Marine Science, 71: 918–924.

Staaterman, E., Paris, C. B., and Helgers, J. 2012. Orientation behavior in fish larvae: a missing piece to Hjort 's critical period hypothesis. Journal of Theoretical Biology, 304: 188–196.

Sudret, B. 2014. Polynomial chaos expansions and stochastic finite element methods. Risk and Reliability in Geotechnical Engineering, pp. 265–300. Ed. by K.-K. Phoon, and J. Ching, Chapter 6. CRC Press. Singapore

Swearer, S. E., Treml, E. A., and Shima, J. S. 2019. A review of biophysical models of marine larval dispersal. *In* Oceanography and Marine Biology: An Annual Review, 57, pp. 325–356. Ed. by S. J. Hawkins, A. L. Allcock, A. E. Bates, L. B. Firth, I. P. Smith, S. E. Swearer, and P. A. Todd, Taylor & Francis Group. London

Torrado, H., Mourre, B., Raventos, N., Carreras, C., Tintoré, J., Pascual, M., and Macpherson, E. 2021. Impact of individual early life traits in larval dispersal: a multispecies approach using backtracking models. Progress in Oceanography, 192:102518 .

Treml, E. A., Roberts, J. J., Chao, Y., Halpin, P. N., Possingham, H. P., and Riginos, C. 2012. Reproductive output and duration of the pelagic larval stage determine seascape-wide connectivity of marine populations. Integrative and Comparative Biology, 52: 525–537.

Treml, E. A., Roberts, J., Halpin, P. N., Possingham, H. P., and Riginos, C. 2015. The emergent geography of biophysical dispersal barriers across the Indo-West Pacific. Diversity and Distributions, 21: 465–476.

Truelove, N. K., Kough, A. S., Behringer, D. C., Paris, C. B., Box, S. J., Preziosi, R. F., and Butler, M. J. 2017. Biophysical connectivity explains population genetic structure in a highly dispersive marine species. Coral Reefs, 36: 233–244.

van Sebille, E., Griffies, S. M., Abernathey, R., Adams, T. P., Berloff, P., Biastoch, A., Blanke, B. *et al.*. 2018. Lagrangian ocean analysis: fundamentals and practices. Ocean Modelling, 121: 49–75.

Vaz, A. C., Paris, C. B., Olascoaga, M. J., Kourafalou, V. H., Kang, H., and Reed, J. K. 2016. The perfect storm: match-mismatch of biophysical events drives larval reef fish connectivity between Pulley Ridge mesophotic reef and the Florida Keys. Continental Shelf Research, 125: 136–146.

Wright, K. J., Higgs, D. M., Belanger, a. J., and Leis, J. M. 2005. Auditory and olfactory abilities of pre-settlement larvae and post-settlement juveniles of a coral reef damselfish (Pisces: Pomacentridae). Marine Biology, 147: 1425–1434.

Appendix A:

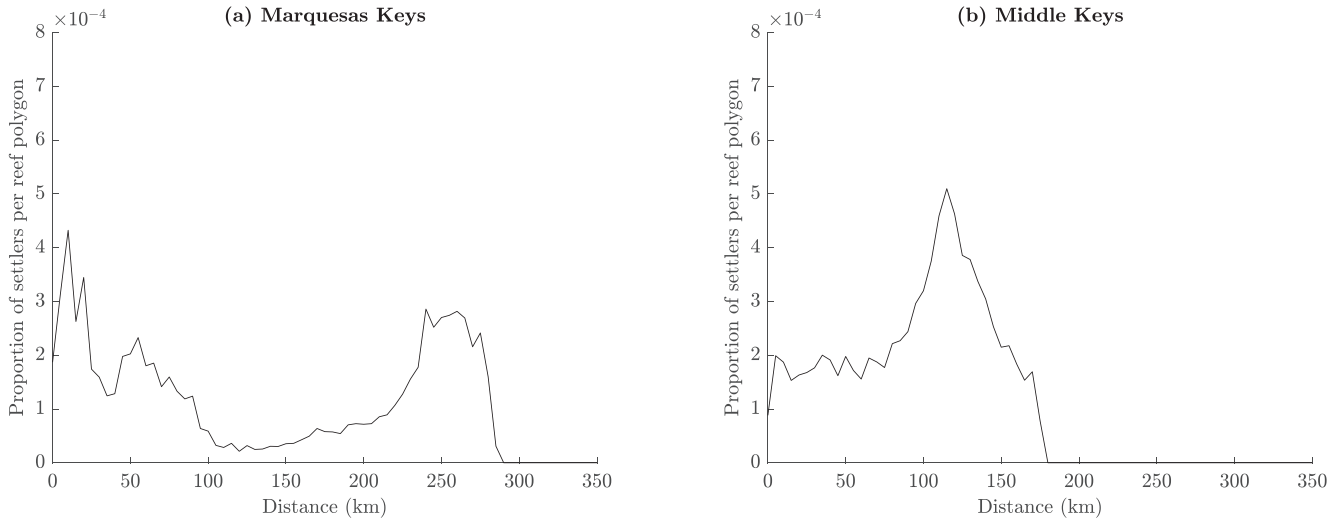

Figure A1. Dispersal kernels from the MK (left panel) and the MidK (right panel). The dispersal kernels are estimated with the CMS using the average values for the input parameters.

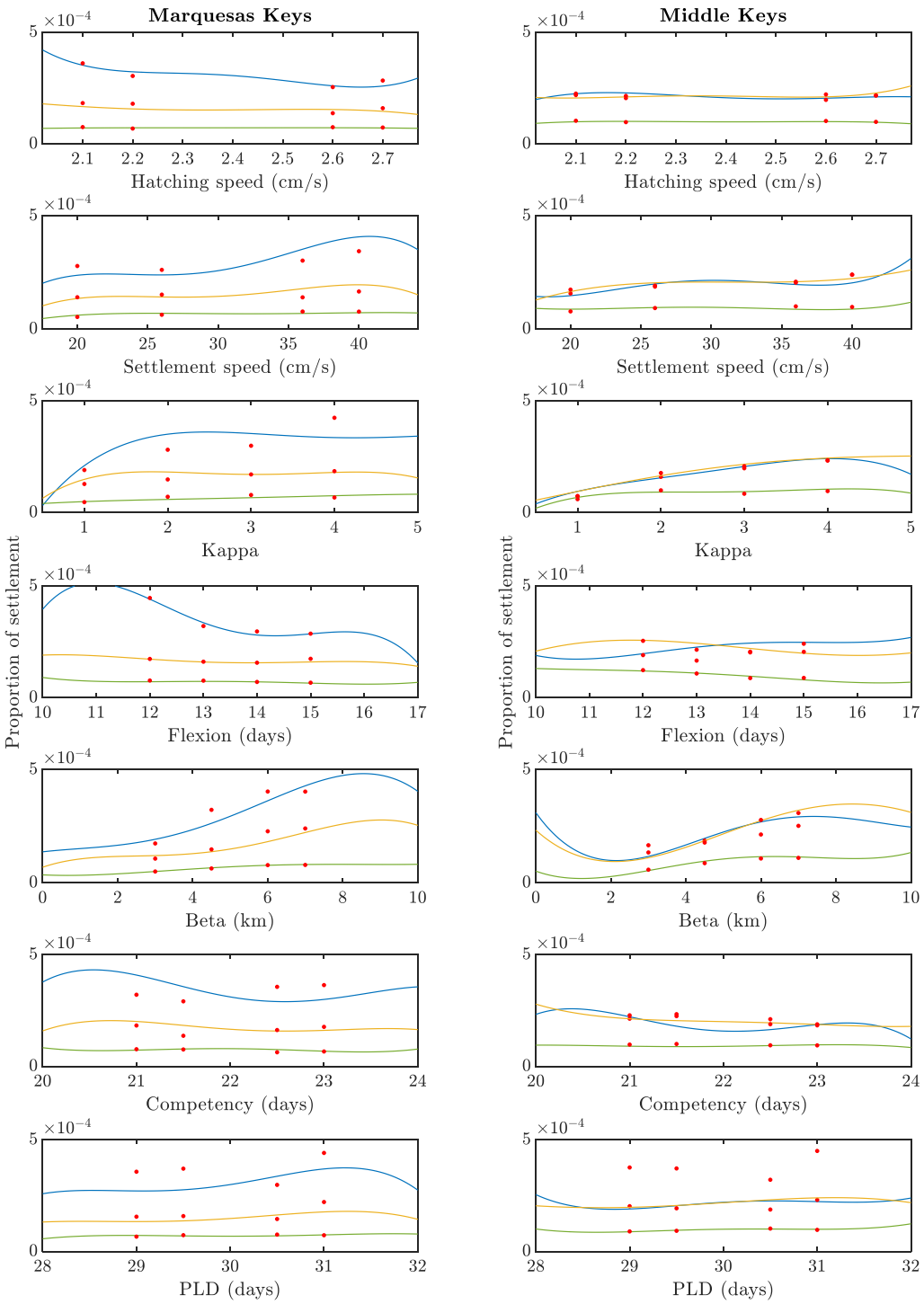

Figure A2. Variation of the proportion of settlement as predicted by the PC surrogate within the seven input parameter ranges for short distance dispersal (0–10 km: blue), medium distance dispersal (10–100 km: orange), and long distance dispersal (> 100 km: green). Results are presented for larvae released from the MK (left) and the MidK (right). Input parameters are varied one-at-a-time, with the values of the other non-varied parameters fixed to their mean. Red points show the proportion of settlement as predicted by 28 validation runs of the CMS.

Table A1. Output variance of the proportion of short distance settlement estimated for each of the seven input parameters tested in the uni-dimensional analysis. The smallest variance per reef is highlighted in bold. Swimming speed at hatching and PLD are the two variables producing the least variance for three out of four reefs considered.

	Sw_{hatch}	Sw_{settle}	κ	*Flexion*	β	*Competency*	*PLD*
DT	1.21e-08	**2.56e-09**	1.42e-07	2.77e-07	3.99e-07	**5.56e-09**	2.61e-08
LK	**6.32e-10**	4.39e-09	1.34e-08	2.69e-09	1.86e-08	4.62e-09	**1.07e-09**
Marquesas	**1.42e-09**	4.49e-09	5.28e-09	9.01e-09	1.57e-08	2.41e-09	**1.54e-09**
MidK	**8.70e-11**	1.03e-09	3.09e-09	9.21e-10	4.87e-09	1.19e-09	**2.12e-10**

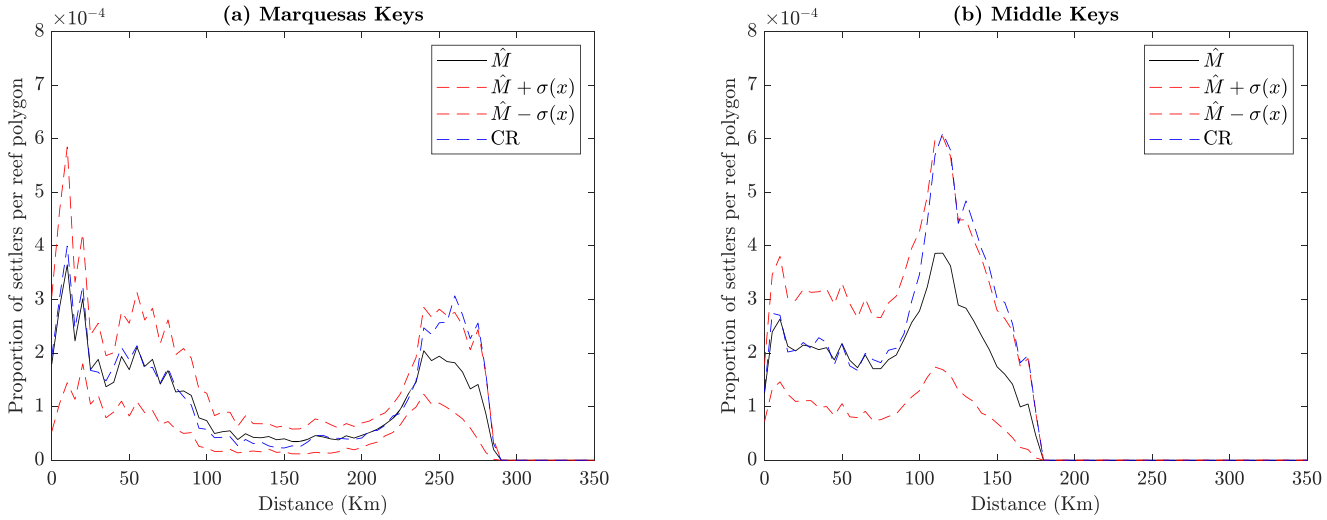

Figure A3. PC estimates of the dispersal kernels for the MK (left panel) and the MidK (right panel). The PC surrogate gives a statistical representation of the uncertainties brought by the biological inputs. Dispersal kernels for both reefs have similar shapes as the CMS control run estimate (CR: blue dotted line). Average proportion of settlement is lower than control run estimate across most dispersal distances, highlighting an overestimation in the control scenario. Standard deviations for both reefs match the peaks of the distributions, and show larger impact of larval behaviours in these regions.

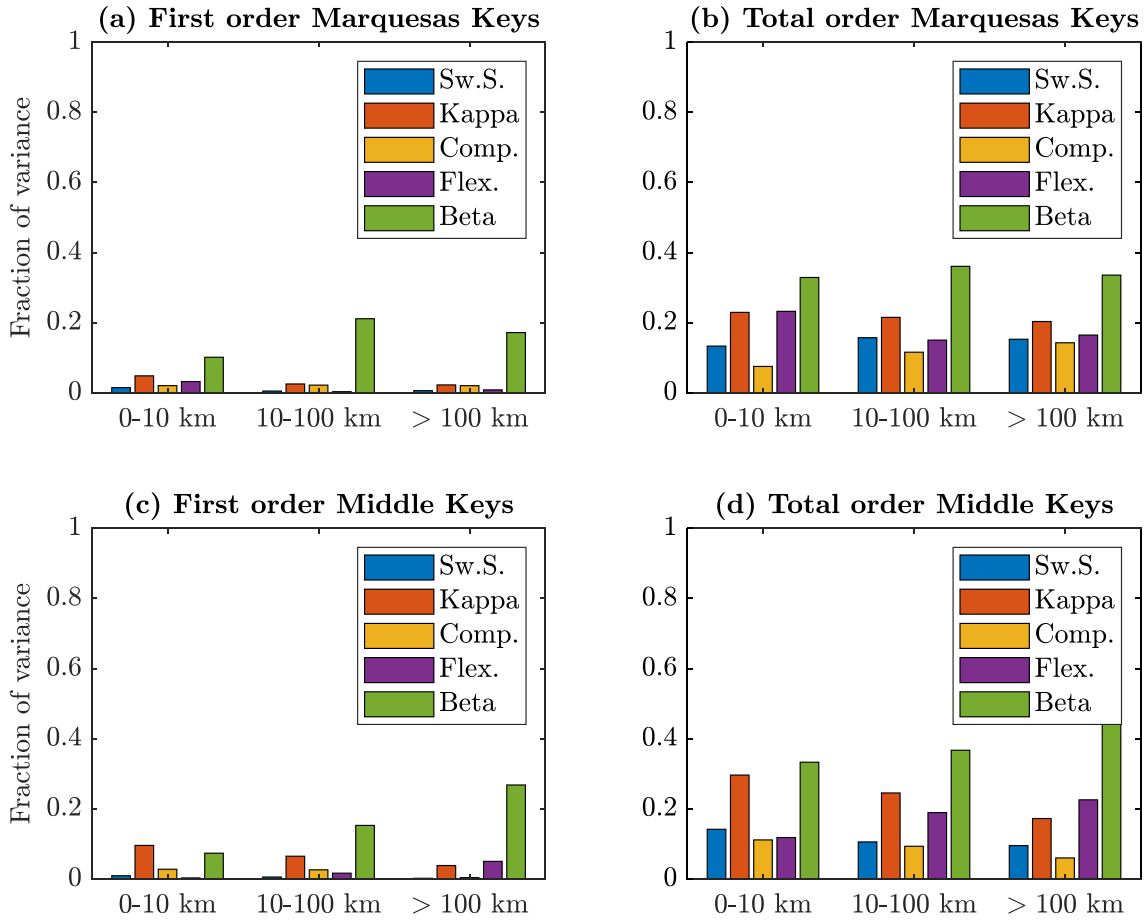

Figure A4. First order ($S_{1...5}$, left) and total order sensitivity indices ($T_{1...5}$, right) for the input parameters Settlement swimming speed, Kappa, Competency, Flexion, and Beta. Variances induced by the parameters are presented for the MK (Top) and the MidK (Bottom) at three dispersal ranges: short (0–10 km), medium (10–100 km), and long (> 100 km).

Appendix B: surrogate validation

The reliability of the uncertainty analysis presented here hinges on the ability of the surrogate to substitute for actual model runs. We, therefore, perform validation tests to gauge the surrogate's accuracy. These tests consist of comparing surrogate and model predictions at a set of N_{vp} additional val-

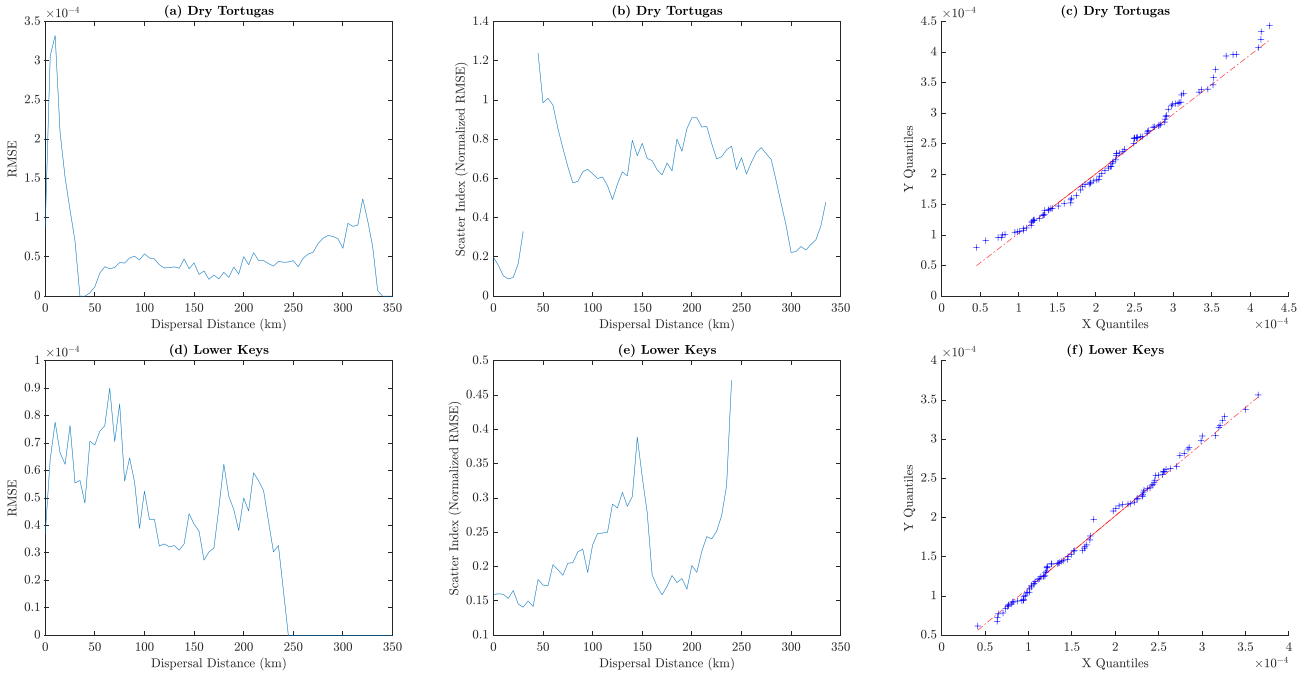

Figure B1. Analysis of the PC surrogate fit with the CMS estimates. Root Mean Square Errors (RMSE) between the surrogate estimates and the CMS predictions for 100 validation points are analyzed along the dispersal kernels for (a) DT and (d) the LK. RMSE are about one order of magnitude lower than the predicted proportion of settlement for all dispersal distances. (b) and (e) Normalization of the RMSE produces the scatter index for both reefs shows increased precision of the surrogate for dispersal distances with a large proportion of settlement and lower precision for dispersal distances with low proportion of settlement. (c) and (f) The quantile to quantile plots show the quantiles of the CMS validation points (*x*-axis) against the quantiles of the surrogate (*y*-axis). The surrogate estimates are close to the CMS validation points distribution (the diagonal line representing a perfect fit). Only a few points on both extreme quantiles seem to be overestimated by the surrogate for DT (c).

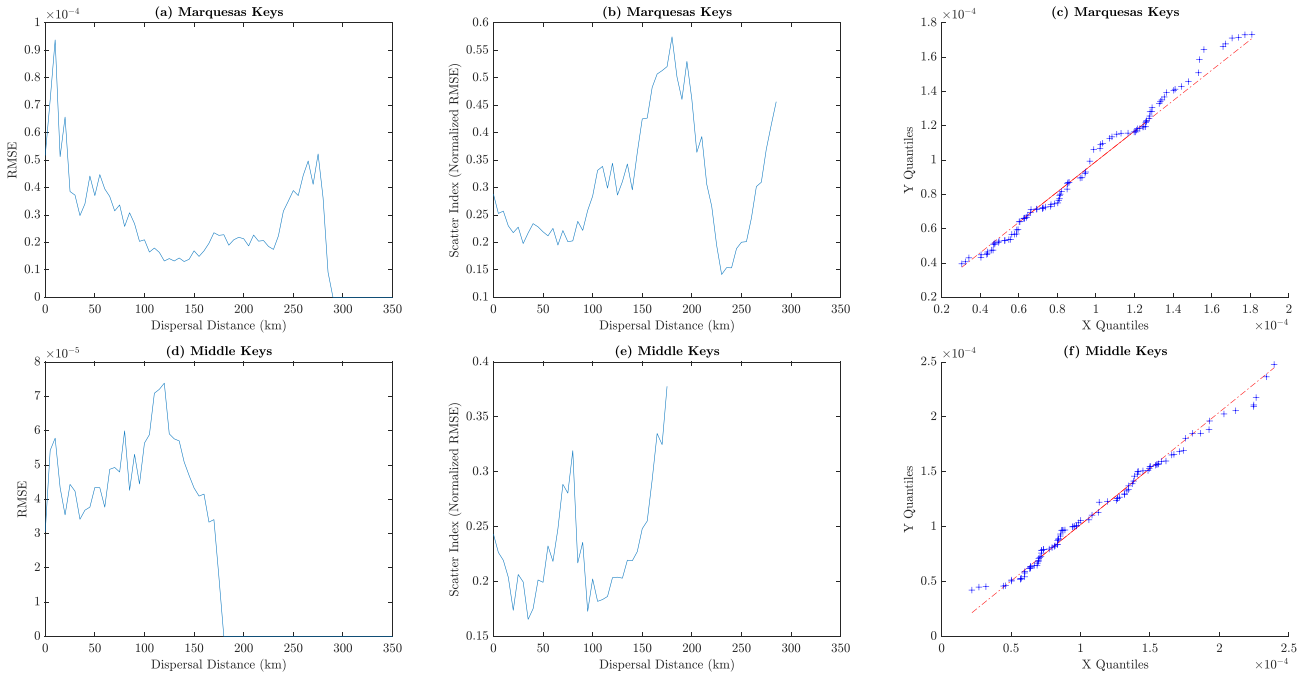

Figure B2. Analysis of the PC surrogate fit with the CMS estimates. Root Mean Square Errors (RMSE) between the surrogate estimates and the CMS predictions for 100 validation points are analysed along the dispersal kernels for the (a) MK and (d) the MidK. RMSE are about one order of magnitude lower than the predicted proportion of settlement for all dispersal distances. (b) and (e) Normalization of the RMSE produces the scatter index for both reefs shows increased precision of the surrogate for dispersal distances with a large proportion of settlement and lower precision for dispersal distances with low proportion of settlement. (c) and (f) The quantile to quantile plots show the quantiles of the CMS validation points (*x*-axis) against the quantiles of the surrogate (*y*-axis). The surrogate estimates are close to the CMS validation points distribution (the diagonal line representing a perfect fit). Only a few points on both extreme quantiles seem to be overestimated by the surrogate for MK (c).

idation sampling points $\{\boldsymbol{\xi}_i^{\mathrm{v}}\}_{1\leq i\leq N_{vp}}$ that are distinct from the training samples. The surrogate prediction error is quantified using the Root Mean Square Errors (RMSE):

$$\mathrm{RMSE} = \sqrt{\frac{1}{N_{vp}}\sum\left(f(\boldsymbol{\xi}_i^{\mathrm{v}}) - f_P(\boldsymbol{\xi}_i^{\mathrm{v}})\right)^2}. \quad \text{(B1)}$$

In addition, a quantile to quantile plot is used to determine if the surrogate and the CMS estimates follow a similar distribution.

The RMSE, computed based on 100 random validation points, follow closely the shape of the dispersal kernels both for DT and the LK (see Figure B1a and d). For both reefs, the RMSE seems to be larger at low values of settlement. This diagnostic is confirmed by the scatter index, the RMSE normalized by the mean of proportion of settlement per dispersal distance (Figure B1b and e). The scatter index is the inverse of the RMSE. The surrogate, therefore, estimates better the high values of settlement compared to low values of settlement. Analysis of the quantile to quantile plots (qqplots: Figure B1c and f) shows that both the surrogate and the CMS estimates share a similar distribution, with only a few points, for DT, at the extremes of the distribution that are underestimated by the surrogate (Y quantile on the plot). These diagnostics show a particularly good estimation of the PC surrogate. The RMSE and sensitivity analysis for the reefs of the MK and the MidK (Figure B2) show that the fit of the PC surrogate is better in the case of the MidK than for the MK.

Appendix C: description of the orientation module

The files described here (available at https://github.com/RomainChaput/connectivity-modeling-system) are an addition to the open source version of the CMS (Paris *et al.*, 2013b, available at https://github.com/beatrixparis/connectivity-modeling-system). All the modifications made to the original CMS code are explained and commented in the code. All references used to develop the code are cited in the code itself.

The orientation module can simulate one out of five different orientation scenarios:

(a) Direct orientation, where the larvae swim towards the nearest reef. The algorithms for this mechanisms have been developed by Codling *et al.* (2004) and previously integrated in the CMS by Staaterman *et al.* (2012). Similar orientation mechanisms have been developed and used by Painter and Hillen (2015). Larval orientation according to this model depends on the direction and distance of the nearest reef and the previous larval heading. When a larva detects a nearby reef, it orients itself imperfectly towards the reef. Therefore, the orientation accuracy is simulated by randomly picking the larva heading from a statistical distribution.

(b) Rheotaxis orientation, where the larvae swim against the local current at a speed slower or equal to that of the currents. This behaviour has been observed in multiple fish larvae species (Kingsford *et al.*, 2002; Leis, 2006; Olszewski *et al.*, 2012). With this behaviour, the larvae are made to swim against a known current. The accuracy of the orientation of the larvae is computed similarly as the previous behaviour.

(c) Cardinal orientation, where the larvae use a magnetic, internal compass to swim towards a pre-determined heading (Bottesch *et al.*, 2016; Cresci, 2020). With this behaviour, the larvae go in a pre-determined direction during orientation. The accuracy of orientation is computed similarly as in direct orientation.

(d) A first continuous orientation scenario, where the larvae swim first against the currents and switch to Direct Orientation when they approach the reefs (Bottesch *et al.*, 2016). The user controls the threshold distance at which this change in behaviours occurs. This orientation scenario combines the two orientation mechanisms detailed in *a* and *b*.

(e) A second continuous orientation scenario, where the larvae orient with Cardinal Orientation first and switch to Direct orientation when they approach the reefs (Cresci, 2020). This orientation scenario combines the two orientation mechanisms detailed in *a* and *c*.

Orientation module added to the open source CMS

'mod_orientation.f90': this module was developed to simulate orientation behaviours in the CMS. It is organized as follow:

- Subroutine 'load_center_data': loads the center of the reefs (used in the behaviour *a*. Direct orientation).
- Subroutine 'nearest_reef': finds the nearest reef and compute the distance between the larva and the reef (used in the behaviour *a*. Direct orientation).
- Subroutine 'calc_orient': computes the heading of the larva relative to the nearest reef and its swimming speed (used in the behaviour *a*. Direct orientation).
- Subroutine 'calc_rheotaxis': computes the heading and swimming speed of the larva relative to the currents (used in the behaviour *b*. Rheotaxis).
- Subroutine 'calc_cardinal': computes the heading using a pre-determined direction and the swimming speed of the larva (used in the behaviour *c*. Cardinal orientation).
- Function 'Harversine': formula to compute angles from latitude and longitude data.

Modifications of the open source CMS

- 'def_globalvariables.f90': defines the type of the variables used in the CMS. This file was modified from its original version to include the variables used in the different orientation behaviours.
- 'input.f90': defines the parameters that the user can input to the model. This file was modified to include the input parameters used for the different orientation behaviours.
- 'loop.f90': computes the new position of the particles. It was modified to include the orientation module.
- 'Makefile' : executable file to compile the CMS. It was modified to include the orientation module.
- 'mod_random.f90': contains different statistical distributions used by the CMS. A code to compute the Von Mises distribution was added to the original file.
- 'move.f90': computes the displacement of the larvae. It was modified to include the different orientation behaviours:
  - Line 240: if the option Mix_orient (*d* and *e*. continuous orientation) is chosen by the user, the program calls the subroutines 'nearest_reef', 'calc_cardinal', or 'calc_rheotaxis', and 'calc_orient'.

  - Line 255: if the option Orient (*a*. Direct orientation) is chosen by the user, the program calls the subroutines 'nearest_reef', and 'calc_orient'.
  - Line 265: if the option Rheotaxis (*b*. Rheotaxis orientation) is chosen by the user, the program calls the subroutine 'calc_rheotaxis'.
  - Line 274: if the option Cardinal (*c*. Cardinal orientation) is chosen by the user, the program calls the subroutine 'calc_cardinal'.
- The type of orientation mechanism used by the larvae during dispersal can be chosen and parametrized by the user in the 'ibm.list' file.

Handling Editor: David M Kaplan