Zigzag persistence for coral reef resilience using a stochastic spatial model

R. A. McDonald†, R. Neuhausler†, M. Robinson, L. G. Larsen, H. A. Harrington, and M. Bruna

A complex interplay between species governs the evolution of spatial patterns in ecology. An open problem in the biological sciences is characterizing spatio-temporal data and understanding how changes at the local scale affect global dynamics/behaviour. Here, we extend a well-studied temporal mathematical model of coral reef dynamics to include stochastic and spatial interactions and generate data to study different ecological scenarios. We present descriptors to characterize patterns in heterogeneous spatio-temporal data surpassing spatially averaged measures. We apply these descriptors to simulated coral data and demonstrate the utility of two topological data analysis techniques—persistent homology and zigzag persistence—for characterizing mechanisms of reef resilience. We show that the introduction of local competition between species leads to the appearance of coral clusters in the reef. We use our analyses to distinguish temporal dynamics stemming from different initial configurations of coral, showing that the neighbourhood composition of coral sites determines their long-term survival. Using zigzag persistence, we determine which spatial configurations protect coral from extinction in different environments. Finally, we apply this toolkit of multiscale methods to empirical coral reef data, which distinguish spatio-temporal reef dynamics in different locations, and demonstrate the applicability to a range of datasets.

1. Introduction

Spatial patterns arise in many natural systems, from systems of chemical species or morphogens [1], cells in developing embryos [2,3], skin patterns on fish and mammals [4,5], and coral colonies in coral reefs [6]. Alan Turing explained the mechanisms behind the spatial patterns observed in morphogenesis—the interplay of diffusion and reactions [7]. Recent work has shed light on the importance of early segregation, or spatial patterning in embryos for successful development [8]. A common denominator of such systems is their complexity: they are dynamic, involve large numbers of particles or agents (e.g., molecules, cells, animals), and are inherently noisy. To elucidate the role of spatial patterns in such systems’ function and spatial evolution requires quantitative tools that can cope with such complexity. In recent years, the area of topological data analysis (TDA) has blossomed to offer multiple promising methods [9]. TDA can provide multiscale summaries of complex data. Here, we dive into the mechanisms of spatial patterning with TDA and other topological descriptors, with shallow-water coral reefs as a case study.

Coral reefs provide a tremendous range of ecosystem services, including biodiversity, fishing, and tourism [10,11]. Due to the complexity and stability of their...
calcium carbonate structure, specifically in shallow waters, coral reefs supply the optimal foundation for various photo-synthetic benthic organisms to settle and grow upon. As such, there is relentless competition for space on the reef. Under human or natural disturbances—which are becoming ever more common with climate change and coastal development—coral reefs have been observed to shift from coral- to algae-dominated states [12]. Many mechanistic models have been proposed for hypothesis-testing about mechanisms that drive spatial patterning and long-term resilience [13–18]. In particular, Mumby, Hastings and Edwards (MHE) developed a simple three-species temporal model for Caribbean coral reefs accounting for corals and two types of algae, algal turfs and macroalgae [19]. The MHE model demonstrated that there are coral- and algae-dominated alternative states and established critical thresholds of fish grazing and coral cover delineating the resilience of each state.

A natural question is how coral resilience is affected by the spatial distribution of coral and other species within the reef. To address this question, we build on the MHE model [19] to develop a stochastic and spatial lattice-based model (sMHE) of a coral reef. Stochasticity is essential to model the unpredictable disturbances that affect coral reefs and enables transitions between the alternative stable states identified by the MHE model. Spatial models, both deterministic (based on partial differential equations) and stochastic (e.g. agent-based and cellular automata models), have been used to study pattern formation [4,5,20]. Due to stochasticity, multiple realizations of the sMHE model are required to produce statistical results and develop insight into the reef-level dynamics while retaining the spatial information. To this end, we use topological descriptors suited to averaging over realizations. We first consider neighbourhood descriptors that quantify the clustering of coral throughout the reef and then appeal to TDA.

TDA is a branch of computational mathematics that summarizes the shape of data through topological invariants [21,22]. Persistent homology (PH), a prominent tool in TDA, takes in data and outputs a multiscale topological summary of features, such as connected components, loops and cavities [23]. Depending on the type of data being studied, PH offers a flexible suite of methodologies which may be adapted to address the research question at hand [24]. PH has previously generated insight into many biological applications [25–27]. While competition for space is known to affect reef dynamics [18], much conventional analysis of reef data focuses on the prevalence of different species (measured through percentage cover). We use TDA to analyse simulated and empirical data, reflecting a square-grid discretization of the seabed domain with 25 × 25 nodes, where each node i is spatially embedded in two dimensions and occupied by either C, M, or T (assuming the seabed area is fully covered). Previous work has extended the MHE model to account for the effect of ocean acidification [41], natural disasters [42] and more complex fishing dynamics [43,44]. Other models consider multiple competing coral and macroalgal types [45] with complex intra-species dynamics [46]. We use the original MHE model as our starting point due to its simplicity and well-understood dynamics.

In the MHE model (2.1), competition between species is represented by the nonlinear terms, which are agnostic of spatial location. One possible modelling framework which would represent the spatial dependence of coral growth would be a system of partial differential equations (PDEs) describing C, M and T as functions of time and space. However, to account for the complexity of the environment, such interspecies interaction should be stochastic, which would not be reflected in a deterministic PDE. Furthermore, we wish to construct a model which simulates data of the same form as photographed coral reefs overlayed with a grid mesh as in figure 1a.

In the stochastic spatial MHE model (sMHE model), we make inter-species competition location-dependent by considering a square-grid discretization of the seabed domain with 25 × 25 nodes, where each node i is spatially embedded in two dimensions and occupied by either C, M, or T (see figure 1b, bottom). The number of nodes (625) was chosen to balance simulation time (which increases exponentially with the number of nodes) with the effectiveness of TDA computations (which distinguish more spatial behaviour at finer resolutions). We introduce a local neighbourhood of radius ε within which interactions can occur (see figure 1b, top). For example, the first term in (2.1a) describes the interaction
between C and T, namely the recruitment of coral through the overgrowth of turf at a rate r. In the sMHE model, this interaction can only occur if there is a node i with turf (for coral to overgrow) within a radius \( \ell \) of a node j with coral, that is, \( T_i = 1, C_j = 1 \) and \(|i - j| \leq \ell\). The interaction is then represented as the ‘reaction’ \( T_i + C_j \rightarrow C_i + C_j \) with rate r (meaning that the transition \( T_i \rightarrow C_i \) occurs with probability \( r\Delta t \) within a time interval \( \Delta t \)). The full set of reactions of the sMHE is

\[
T_i + C_j \rightarrow C_i + C_j, \quad C_i + M_j \rightarrow T_i, \quad C_j + M_i \rightarrow T_i, \quad M_i + M_j \rightarrow T_i + T_i
\]

(2.2a)

and

\[
M_i + M_j \rightarrow T_i, \quad C_j + M_i \rightarrow T_i, \quad C_i + M_j \rightarrow T_i + T_i
\]

(2.2b)

where \( \nu_1, \nu_2 \) are neighbourhood-dependent functions (see electronic supplementary material, §1, for the full specification of equations (2.2)). Throughout this work, we keep all the model parameters fixed at \( r = 1, \ d = 0.4, \ a = 0.2, \ \gamma = 0.75 \) except the neighbourhood radius \( \ell \) and the grazing rate \( g \) at which fish graze on macroalgae. The fixed parameters are taken from [19]. We initialize the model to initial global densities \( C(0), T(0), M(0) \), using either a random initial configuration (uniformly distributed) or a coral-cluster initial configuration (where all coral nodes are placed together in the centre of the domain, and \( M \) and \( T \) are uniformly distributed around in the remaining space; see figure 1b, bottom). The sMHE model (2.2) is simulated as follows: at fixed time steps \( \Delta t \), each node i is considered in turn and is allowed to react with a probability according to the neighbourhood-dependent rates (see electronic supplementary material, §1, for details on the stochastic simulation algorithm). A simulation of the sMHE model is a sequence of ternary matrices (see figure 1c, bottom), which we call snapshots, in which each entry indicates which of the three species occupies that location within the reef. From the matrix, we can extract the global fractional cover of each species at any time, e.g. \( C(t) = \sum_i C_i(t)/625 \).

3. Descriptors for coral data analysis

A simple, effectively non-spatial descriptor is the fractional cover, given by the proportion of nodes of a given type in the reef (figures 1c and 2a). We explore the spatial dynamics of the sMHE model using a collection of neighbourhood descriptors, PH and zigzag persistence.

3.1. Neighbourhood descriptors

We introduce nine neighbourhood descriptors to quantify the average neighbourhood composition of nodes in the reef. Let node \( i \) be of type C (\( C_i = 1 \)). The neighbours of node \( i \) are those nodes within a radius \( \ell \) of \( i \) (figure 1b, middle). We count the number of neighbours \( \Delta r \) of node \( i \) that are of type \( \sigma \) for \( \sigma \in \{ C, T, M \} \). We have \( \Delta r = \{ 0, \ldots, \Delta r \} \), where \( \Delta r \) is the total number of neighbours of node \( i \) (e.g. \( \Delta r = 8 \) for an internal node). Then we define the neighbourhood descriptors \( \sigma_i \), as the average of \( \Delta r/\Delta r \) across all nodes \( i \) of type C. The descriptor \( \sigma_{C_i} \) for instance, gives the average fraction of coral neighbours that are coral (figure 2e). The other six descriptors are defined similarly, considering nodes \( i \) of type T or M to give nine descriptors in total (see electronic supplementary material, §1, for details). We may average the fractional coral cover over many realizations (figure 1c) to give a non-spatial summary of the model’s behaviour. However, the same coral fractional cover can take many different spatial patterns. For
example, the random or coral-cluster initial configurations can be set with the same fractional cover (0.33), yet the local neighbourhood information differs significantly ($C_C$ (random) < 0.2 whereas $C_C$ (coral-cluster) > 0.8) (figure 1b, bottom). This local neighbourhood descriptor highlights that coral occupies more than 80% of the neighbours of coral nodes in the coral-cluster configuration but less than 20% for the random configuration, and therefore distinguishes spatially inhomogeneous reefs with the same fractional covers.

### 3.2. Persistence

PH offers an algorithmic way to quantify the connectivity of multiscale data. We represent a snapshot of the sMHE model by a sequence of cubical complexes. A cubical complex is a data structure that represents nodes on a grid by vertices and connects adjacent vertices with edges and squares. To encode information about the density of coral nodes, we assign an integer to each vertex $i$ with $C_i = 1$ based on the number of direct coral neighbours. Mathematically, we define a density filtration function $f: \mathcal{I} \rightarrow N$, where $\mathcal{I}$ is the set of nodes in the reef and $N = \{0, ..., 8\}$ [47,48]. We use cubical complexes and $f$ to create a multiscale lens called a filtration.

In the first filtration step, we create a cubical complex using only those coral nodes with eight direct coral neighbours (i.e. completely surrounded by coral nodes), adding edges and squares between adjacent coral nodes. At each subsequent step, we include coral nodes with 7, 6, ... coral neighbours to the cubical complex. We define a nested sequence of cubical complexes according to the number of direct coral neighbours, $K_8 \subset K_7 \subset \ldots \subset K_2 \subset K_1$.

For a given time-snapshot of the sMHE model, we build the filtration and then compute standard PH (see electronic supplementary material, S2, for full details). PH quantifies the topological features, such as clusters $H_0$ (i.e. connected components) or loops $H_1$ (i.e. one-dimensional holes) across the filtration. The appearance and disappearance of components and loops across the filtration can be visualized as a multiset of intervals called a barcode. Here, barcodes quantify features according to their size since large numbers of direct coral neighbours indicate large clusters (figure 2c,d). In figure 2d, the barcodes capture the temporal evolution of the data, from the random spatial structure described by many short bars at $t=0$ to the single long bar at time $t=100$ representing one large coral cluster.

PH considers specific times of the sMHE data independently, making it difficult to decide whether a single component of coral persists or whether different coral clusters appear at each timestep. We want to trace the time evolution of specific spatial features in the sMHE model. Due to the non-monotonicity of coral dynamics (i.e. coral locations can be occupied by other species and then return to coral), standard PH is not suitable, since coral clusters are not nested over time.

### 3.3. Zigzag persistence

To track the evolution of topological features over many time steps, we propose to use zigzag persistence [35], which generalizes the notion of filtration to a zigzag diagram. At each time point $t_w$, we choose one of the complexes from the filtration described above to represent a snapshot of the
sMHE model. We choose $K_{n}^{l}$, the cubical complex obtained by including coral nodes with at least one coral neighbour. We then insert the intersection of every pair of successive complexes into this sequence, giving

$$K_{n}^{l} \supset (K_{n}^{l} \cap K_{n-1}^{l}) \supset K_{n-1}^{l} \supset (K_{n-1}^{l} \cap K_{n-2}^{l}) \supset \cdots \supset K_{1}^{l}. \quad (3.1)$$

A diagrammatic representation of (3.1) with three time points is given in figure 2e. We use this sequence to compute connected components (i.e. $H_{0}$ of (3.1)), which emerge and disappear over many timesteps in a simulation of the sMHE model. We perform a pre-processing step to reduce the noise of the sMHE data before the computation of zigzag persistence (see electronic supplementary material, §3).

Due to the stochastic nature of the sMHE model, the zigzag barcode of a single simulation would not provide an accurate summary of the model's behaviour. Persistence landscapes offer a way to average the spatial information encoded in barcodes [28]. We create a persistence landscape from a simulation of the sMHE model as follows. For each component of coral that is detected in a simulation of the model, we plot a landscape peak at $(t_{l}, s_{l})$, where $t_{l}$ is halfway through the component's life span, and $s_{l}$ is half of the length of time for which it exists. For example, if a component is born at time $t = 50$ and dies at time $t = 100$, a landscape peak is placed at $(75, 25)$. A 'tent' is then constructed by connecting $(t_{l}, s_{l})$ linearly down to the birth and death times on the $t$-axis. The $k$th persistence landscape $\lambda_{k}$ takes, at every time, the $k$th largest value among all the 'tents' (see electronic supplementary material, §3, for a full explanation of this process). The landscape $\lambda_{4}$ is the average of $\lambda_{4}$ over many simulations, and the maximum of each $\lambda_{k}$ may be interpreted as the average half-lifetime of the $k$th most persistent cluster. Therefore, zigzag persistence landscapes quantify the typical spatial information across multiple simulations of the sMHE model and rank which features are most significant over time.

The first landscape ($\lambda_{1}$) in figure 2f highlights that a single cluster of coral dominates the domain from $t = 0$ through $t = 100$. We can observe that the time-persistence of the second and third landscapes ($\lambda_{2}, \lambda_{3}$) decreases as time increases, suggesting that smaller components either disappear or join with the main one as time increases. Zigzag persistence, therefore, confirms that, on average, a single component is emerging over time, which PH alone cannot conclude.

We use these tools to explore the model dynamics and spatial structure. Specifically, we consider the effect of local neighbourhood radius $\ell$, the initial configuration of species (random or coral-cluster configurations), and the rate $g$ that fish graze on macroalgae, and quantify how changing these three parameters affects the sMHE model.

4. Results

4.1. Spatial interactions lead to coral clustering

The range of the spatial interactions in the sMHE model is controlled by the neighbourhood radius $\ell$ (figure 1b). For large values of $\ell$, the neighbourhood of interaction spans the whole domain, and the sMHE model reduces to a non-spatial stochastic model. However for small values of $\ell$ the rates of reactions (2.2) are highly dependent on a node's immediate neighbours. The neighbourhood descriptor $CC$ and PH describe the spatial pattern observed in simulations for different values of the neighbourhood radius $\ell$ (see one realization in figure 3). In particular, larger clusters of coral appear when the interaction range $\ell$ is small, whereas we observe little spatial patterning when $\ell$ is large. Zigzag persistence determines over multiple timesteps that a stable cluster persists over time in the sMHE model with $\ell = 1.45$, whereas no such cluster persists in the non-spatial model (see electronic supplementary material, figure S9).

4.2. Zigzag persistence distinguishes initial configurations

Intuitively, we may think that a single large coral cluster might offer the best conditions for coral resilience over time. To test this hypothesis, we initialize the sMHE model to two initial conditions with identical fractional covers ($C(0), M(0), T(0)$) but very different spatial configurations, namely the random and coral-cluster initial configurations (figure 1b). Perhaps surprisingly, depending on the grazing conditions, the analysis suggests that coral can be more robust over time when initialized with the random configuration, whereas it stagnates or even becomes extinct when initialized with the coral-cluster configuration. The neighbourhood descriptors, as well as the average cluster size and total number of clusters, distinguish the two initial configurations, showing significant differences between the two cases at $t = 0$. However, these descriptors cannot discern the two cases as time progresses (figure 4b). On the other hand, zigzag persistence landscapes $\lambda_{2}, \lambda_{3}$—averaged over many simulations—significantly differ between the two cases over the whole simulation time (figure 4e). We can understand this result by noting that the random initial configuration yields a more significant average number of coral-turf neighbours ($T_{r}(0) = 0.765$ and 0.113 for random and coral-cluster initial conditions, respectively), increasing the locations where coral growth can occur (via the interaction $T_{r} + C_{j} \rightarrow C_{j} + C$) when compared with the coral-cluster initial configuration. In this way, under certain parameter regimes, the spatial configuration of nodes in the sMHE model may determine the long-term behaviour of the system.

4.3. Zigzag persistence describes coral extinction pathway

In many reefs, coral’s spatio-temporal dynamics critically depend on the fish population feeding on the macroalgae [49]. Thriving fish populations (i.e. high grazing rate $g$) keep macroalgae levels low, which leads to better conditions for coral to flourish. By contrast, overfishing and natural disasters shrink the fish population and hence $g$, which may result in macroalgae overgrowth and coral decay.

The MHE model (2.1) captured these coral-reef interactions and established that the grazing rate $g$ is a bifurcation parameter, where low grazing leads to coral extinction, and intermediate grazing drives the system to display two alternative states [19].

The sMHE model (2.2) reproduces this behaviour: for $g < 0.52$ the system evolves to a macroalgae-dominated reef; for $g > 0.55$ the system evolves to a coral-dominated reef. In the region of multistability ($g \in [0.52, 0.55]$), we observe that simulations can evolve to either the coral- or macroalgae-dominated state. However, in contrast to original MHE model, where the long-term behaviours of simulations in the metastable region are determined by their initial condition, the stochastic nature of the sMHE model means that identical
initial conditions may lead to different outcomes. At $g = 0.53$, we observed that around half of all simulations (initiated with equal numbers of coral, turf and macroalgae nodes in the random configuration) evolved to the macroalgae-dominated state, with the other half converging to a coral-dominated state.

Metastability implies that the system takes a long time to converge to such a state (up to 1000 timesteps in some realizations). We explore whether summaries of the early species’ behaviour (we choose $t \in [0, 100]$) predict the system outcome. Non-spatial descriptors (e.g. the fractional covers $C(t), M(t)$; see electronic supplementary material, figure

---

**Figure 3.** Effect of neighbourhood threshold spatial and non-spatial versions of the sMHE model are compared by running simulations up to time $t = 10$. The grazing rate is kept constant at $g = 0.53$, the random initial configuration is used with equal initial fractional covers $C(0), T(0)$ and $M(0)$. The neighbourhood size is varied: $\ell = 1.45, 2.9, 4.3, 36$. As described in figure 1, a neighbourhood size of $\ell = 1.45$ gives a spatial model, where only the immediate eight neighbours of each node affect its transition as the model is updated. The values $\ell = 2.9, 4.3$ give spatial models where a larger grid of nodes affects the reaction rates. When $\ell = 36$, all nodes are considered neighbours of all others, giving a non-spatial model. The snapshots of the sMHE model are printed at time $t = 0$ and at time $t = 10$ for each of these scenarios. The coral–coral neighbourhood descriptor, $C_C$, distinguishes the three cases. The value $C_C = 0.559$ for the $\ell = 1.45$ spatial model indicates that coral clusters more tightly together in this case, whereas the value $C_C = 0.248$ shows that this does not happen for the non-spatial model. The adjacent persistence barcodes give further details of this difference. For the $\ell = 1.45$ spatial model, there are a few large components of coral (indicated by solid bars) and some enclosed loops (indicated by dotted bars). For the non-spatial case, there are many smaller components of coral and no loops. The $\ell = 2.9, 4.3$ plots give intermediate results in both $C_C$ and the persistent barcode.
**Figure 4.** Effect of initial configuration comparison of 100 realizations of the sMHE model initialized with either the random or coral-cluster initial configurations from figure 1b, with grazing fixed to $g = 0.53$ and initial fractional covers $C(0) = M(0) = 0.15$. (a) Average persistence landscapes $\lambda_1$, $\lambda_2$, $\lambda_3$. While the random initial configuration leads to three non-trivial landscapes, the coral-cluster initial configuration gives a significant first landscape $\lambda_1$. These average landscapes indicate that the random initial configuration leads to coral clustering in many different components, which persist as separate patches up to at least time $t = 20$. By contrast, the coral-cluster initial configuration begins with a single connected component of coral. The analysis shows that few other components form throughout the simulation and that coral remains clustered in the initial central patch. Examples of model snapshots at $t = 0$ and $t = 20$ are printed at the side of each landscape plot. (b) Coral fractional cover $C(t)$ and neighbourhood descriptors. The plot of $C(t)$ shows that the random initial configuration favours coral growth, since the fractional coral cover increases in these simulations. By contrast, coral appears to die out when simulations start with the coral-cluster initial configuration. The $T_C$, $M_C$ and $C_C$ plots show that these descriptors clearly distinguish the two initial configurations at time $t = 0$ but cannot tell them apart after a small number of timesteps. Neighbourhood descriptors give similar values for the two initial configurations from time $t = 5$ onwards.
S11) can give an early indication of the system outcome in the metastable region.

But what makes coral die out in some runs of the sMHE in the metastable region and resist in others?

To address this question, we turn to zigzag persistence. Zigzag persistence can predict the system’s outcome (figure 5a) but also characterizes how the coral’s spatial clustering affects coral outcome (figure 5b). In particular, we find that \( \lambda_2 \) and \( \lambda_3 \) (corresponding to the second and third most dominant topological features) are higher in simulations where coral eventually dies out. The difference in landscape size for different outcomes indicates that under these reef
conditions (in particular, low coral-turf neighbours, $T_C(0) = 0.361$), coral extinction occurs through multiple small clusters (as opposed to one cluster that shrinks to nothing), while coral domination establishes itself as one persistent cluster.

4.4. Topology distinguishes empirical data with same coral cover

We next use the longitudinal spatial data from the Moorea IDEA project [50] to exemplify the use of the proposed topological descriptors. Our methods would be ideally suited for a grid with finer spatial resolution, but data of coral reefs via drones [51] which offer higher spatial resolution currently lack multi-year collection. Here, we apply our topological methods to two multi-year datasets from the Rarotonga reef at two different locations (see figure 6, locations A and B). While the percentage cover of coral is similar at the start ($t = 0$) and at the end ($t = 12$) of the 12-year period for both locations, persistence barcodes and zigzag landscapes distinguish the spatio-temporal evolution (see figure 6, landscapes).

The coral configurations in the two locations (A and B) resemble the coral-cluster and random configurations of the sMHE model, respectively. Similar to the model analysis in figure 4, the coral–coral neighbourhood descriptor ($C_C$) is the same for both locations at the end of the time-course. However, the persistence barcodes and zigzag landscapes differ. Barcodes from the empirical coral-cluster (figure 6, location A) have one persistent bar for each time iterate, whereas the dataset analogous to the random configuration (figure 6, location B) has many short bars in the early iterates and one long bar at the final timestep. The spatial–temporal evolution at different locations can be distinguished by the zigzag landscapes (figure 6, landscapes). In location A, the first zigzag landscape is larger than for location B, and the peaks of landscapes for location B occur later than for location A. The topological analysis therefore reveals that, in location A, one large cluster persists across twelve years, whereas in location B, many small clusters join together over time. Our model predicts that configurations similar to location A are more likely to lead to coral extinction than those similar to location B. To justify this prediction, we would require empirical reef data with better resolution (spatial and temporal). In future, we aim to compare data with coral evolution models such as the sMHE model.

5. Discussion

Motivated by ecology and evolution, as well as the increasing availability of spatial data of such processes, we introduced a stochastic spatio-temporal lattice-based model (sMHE model (2.2)) of coral reefs. We collected data through computational experiments and proposed topological descriptors to quantify coral behaviour and predict mechanisms. Specifically, we explored these descriptors on multiple realizations of coral reef dynamics under changes to the initial configurations and the values of key model parameters.

The evolution of the sMHE model depends on two factors. First, species evolve based on the rate values in equations (2.2) (which model internal and external factors affecting the reef as in the original MHE model [19]) and, second, on the spatial arrangement of nodes in the reef. A combination of both factors determines which reactions are possible (depending on the make-up of the neighbourhood) and more likely to occur. The fish grazing rate $g$ is a critical parameter in both the MHE and sMHE models, with macroalgae dominating for low enough $g$ and coral dominating for high enough $g$ (regardless of the spatial arrangement of species in the reef). By contrast, in the intermediate metastable region, the coral- and macroalgae-favoured reactions balance out so that the spatial arrangement becomes a deciding factor. Zigzag persistence can discern these multiple pathways, intricately dependent on species’ competition. For example, we found that the random initial configuration yields higher coral growth than the coral-cluster initial configuration (figure 4) under turf abundance. Conversely, when macroalgae growth overwhelms coral, zigzag persistence suggests that coral goes extinct by becoming dissected into many components (figure 5b). This indicates that, when macroalgae prevalence is high, coral survives better when clustered together, thus limiting the macroalgae-over-taking-coral interaction in (2.2). Together, these insights show the potential use of our model in helping assess the vulnerability of reefs and better design artificial reefs. By tweaking our model to the conditions of a reef of interest, one can theorize coral’s ideal spacing for survival.

The region of metastability of the sMHE model (in which different outcomes are possible) results from including spatial and stochastic dynamics in our model. This variation in outcomes motivates questions such as: are there disturbances or interventions that could ‘shock’ reefs into one state or another? Could this region make the fate of reefs more ‘reversible’ than one would assume in the analysis of deterministic models? Could this behaviour explain some of the ‘noise’ observed in real data, with high coral mortality in some locations and coral survival in others with similar conditions?

The advantage of topological spatial descriptors is their versatility in analysing spatial and temporal structure in complex data. With the flexibility of PH to propose and adapt different filtrations, the standard PH pipeline can be tailored to study a wide range of other spatially patterned systems. Here we showcased the power of zigzag persistence, a topological measure that has recently benefited from improved computation [40]. We combined zigzag persistence with statistical landscapes [28] to enhance the identification of the geometry of the initial configuration of species as well as the geometric mechanisms of species competition and, ultimately, coral extinction in the metastable parameter region.

Since zigzag persistence quantifies spatial patterns in images, it can be used as a pre-processing step prior to classification. For example, if using data sources such as drones or satellite images, one may vectorize persistence using persistence landscapes, which can then be fed into machine learning algorithms. Based on the positive findings from analysing short-time duration data generated by the sMHE model, a promising future direction is to apply topological descriptors to such high-resolution, short-time duration drone data. Given such highly resolved data, we expect the classification, annotation and localization of species in images with machine learning and topological statistics will become automatic.

We have found that PH and zigzag persistence provide valuable insights into the spatial dynamics of the sMHE model. Topological quantification has been used to describe
complex biological models, with different filtrations and topological descriptors used depending on the specific model in question [25,52,53]. Such topological analyses may, in future, enable the comparison and validation of complex mechanistic models by adapting topological approximate Bayesian computation [54].
References

1. Gray P, Scott SK. 1984 Autocatalytic reactions in the isothermal, continuous stirred tank reactor: oscillations and instabilities in the system A + 2B → 3B; A → C. Chem. Eng. Sci. 39, 1087–1097. (doi:10.1016/0009-2509(84)87017-7)

2. Piqué O. 2003 The segmentation clock: converting embryonic time into spatial pattern. Science 301, 328–330. (doi:10.1126/science.1085887)

3. Gregor T, Bialek W, Van Steveninck RDR, Tank DW, Wieschaus EF. 2005 Diffusion and scaling during early embryonic pattern formation. Proc. Natl Acad. Sci. USA 102, 18403–18407. (doi:10.1073/pnas.0509483102)

4. Volkenring A, Sandstedt B. 2015 Modelling stripe formation in zebrafish: an agent-based approach. J. R. Soc. Interface 12, 20150812. (doi:10.1098/rsif.2015.0812)

5. Murray JD. 2001 Mathematical biology II: spatial models and biomedical applications, vol. 3. New York, NY: Springer.

6. Adjeroud M. 1997 Factors influencing spatial patterns on coral reefs around Moorea, French Polynesia. Mar. Ecol. Progress Ser. 159, 105–119. (doi:10.3354/meps159105)

7. Turing AM. 1952 The chemical basis of morphogenesis. Phil. Trans. R. Soc. Lond. B 237, 37–72. (doi:10.1098/rstb.1952.0012)

8. Warmflash A, Sone B, Etoc F, Sigga ED, Brivanlou AH. 2014 A method to recapitulate early embryonic spatial patterning in human embryonic stem cells. Nat. Methods 11, 847–854. (doi:10.1038/nmeth.3016)

9. Hoef LV, Adams H, King EJ, Ebert-Uphoff I. 2022 A primer on topological data analysis to support image analysis tasks in environmental science. (http://arxiv.org/abs/2207.10552)

10. Moberg F, Folke C. 1999 Ecological goods and services of coral reef ecosystems. Ecol. Econ. 29, 215–233. (doi:10.1016/S0921-8009(99)00009-9)

11. Bellwood DR, Hughes TP, Folke C, Nyström M. 2004 Confronting the coral reef crisis. Nature 429, 827–833. (doi:10.1038/nature02691)

12. Hughes TP. 1994 Catastrophes, phase shifts, and large-scale degradation of a Caribbean coral reef. Science 265, 1547–1551. (doi:10.1126/science.265.5178.1547)

13. Meron E. 2012 Pattern-formation approach to modelling spatially extended ecosystems. Ecol. Model. 234, 70–82. (doi:10.1016/j.ecolmodel.2011.05.015)

14. Gomos KD, Karl SA. 2013 Genetic relatedness does not retain spatial pattern across multiple spatial scales: dispersal and colonization in the coral, Porites lobata. Mar. Ecol. Prog. 472, 3721–3736. (doi:10.1139/mep.2012-0586)

15. Fricke HW, Knauer B. 1986 Diversity and spatial pattern of coral communities in the Red Sea upper twilight zone. Oecologia 71, 29–37. (doi:10.1007/BF00377316)

16. Bradbury RH, Young PC. 1983 Coral interactions and community structure: an analysis of spatial pattern. Mar. Ecol. Prog. Ser. 11, 265–271. (doi:10.3354/meps011265)

17. Maguire LA, Porter JW. 1977 A spatial model of growth and competition strategies in coral communities. Ecol. Model. 3, 249–271. (doi:10.1016/0304-3800(77)90007-2)

18. Mummy PJ, Foster NL, Fahy EAG. 2005 Patch dynamics of coral reef metacommunities under chronic and acute disturbance. Coral Reefs 24, 681–692. (doi:10.1007/s00338-005-0058-5)

19. Mummy PJ, Hastings A, Edwards HJ. 2007 Thresholds and the resilience of Caribbean coral reefs. Nature 450, 98–101. (doi:10.1038/nature06252)

20. Pearson JE. 1993 Complex patterns in a simple system. Science 261, 189–192. (doi:10.1126/science.261.5118.189)

21. Carlsson G. 2009 Topology and data. Bull. Am. Math. Soc. 46, 255–308. (doi:10.1090/S0273-0979-09-01249-X)

22. Edelsbrunner H, Harer JL. 2022 Computational topology: an introduction. Providence, RI: American Mathematical Society.

23. Ghrist R. 2008 Barcodes: the persistent topology of data. Bull. Am. Math. Soc. 45, 61–75. (doi:10.1090/S0273-0979-07-01191-3)

24. Otter N, Porter M, Tillmann U, Grindrod P, Harrington H. 2017 A roadmap for the computation of persistent homology. EPJ Data Sci. 6, 17. (doi:10.1140/epjds/s13688-017-0109-5)

25. McGuill MR, Volkenring A, Sandstedt B. 2020 Topological data analysis of zebrafish patterns. Proc. Natl Acad. Sci. USA 117, 5113–5124. (doi:10.1073/pnas.1917366117)

26. Stolz BJ, Kaeppeler J, Markelc B, Braun F, Lipsmeier F, Muschel RJ, Byrne HM, Harrington HA. 2022 Multiscale topology characterizes dynamic tumor vascular networks. Sci. Adv. 8, eabm2456. (doi:10.1126/sciadv.826456)

27. Taylor D, Klimm F, Harrington HA, Kramar M, Mischaikow K, Porter MA, Mucha PJ. 2015 Topological data analysis of contagion maps for examining spreading processes on networks. Nat. Commun. 6, 7723. (doi:10.1038/ncomms8723)

28. Bubenik P. 2015 Statistical topological data analysis using persistence landscapes. J. Mach. Learn. Res. 16, 77–102.

29. Wasserman L. 2018 Topological data analysis. Annu. Rev. Stat. Appl. 5, 501–532. (doi:10.1146/annurev-statistics-010717-100045)

30. Adams H et al. 2017 Persistence images: a stable vector representation of persistent homology. J. Mach. Learn. Res. 18, 1–35.

31. Topaz CM, Ziegelmeier L, Halverson T. 2015 Topological data analysis of biological aggregation models. PLoS ONE 10, e0126383. (doi:10.1371/journal.pone.0126383)

32. Cohen-Steiner D, Edelsbrunner H, Morozov D. 2006 Vines and vineyards by updating persistence in linear time. In Proc. 22nd Annual Symp. on Computational Geometry, Sedona, AZ, USA, 5–7 June 2006, pp. 119–126. (doi:10.1145/1137856.1137877)
33. Kim W, Mémoli F. 2018 Formigrams: clustering summaries of dynamic data. In 30th Canadian Conf. on Computational Geometry, Winnipeg, Canada, 8–10 August 2018, pp. 180–188.

34. Kim W, Mémoli F. 2021 Spatiotemporal persistent homology for dynamic metric spaces. Discrete Comput. Geom. 66, 831–875. (doi:10.1007/s00454-019-00168-w)

35. Carlsson G, Silva Vde. 2010 Zigzag persistence. Algebraic Geom. Topol. 10, 367–405. (doi:10.2140/agt.2010-02364)

36. Tausz A, Carlsson G. 2011 Applications of zigzag persistence landscapes. (http://arxiv.org/abs/1108.3545)

37. Botnan M, Lesnick M. 2018 Algebraic stability of zigzag persistence modules. Algebraic Geom. Topol. 18, 3133–3204. (doi:10.2140/agt.2018.18.3133)

38. Corcoran P, Jones CB. 2017 Modelling topological features of swarm behaviour in space and time with persistence landscapes. IEEE Access 5, 18534–18544. (doi:10.1109/ACCESS.2017.2749319)

39. Myers A, Khasawneh F, Munch E. 2022 Temporal network analysis using zigzag persistence. (http://arxiv.org/abs/2205.11338)

40. Anthony KRN, Maynard JA, Diaz-Pulido G, Mumby PJ, Marshall PA, Cao L, Hoegh-Guldberg O. 2011 Ocean acidification and warming will lower coral reef resilience. Glob. Change Biol. 17, 1798–1808. (doi:10.1111/j.1365-2486.2010.02364.x)

41. Anthony KRN, Maynard JA, Diaz-Pulido G, Mumby PJ, Marshall PA, Hoegh-Guldberg O. 2011 Ocean acidification and warming will lower coral reef resilience. Glob. Change Biol. 17, 1798–1808. (doi:10.1111/j.1365-2486.2010.02364.x)

42. Blackwood JC, Hastings A, Munday PJ. 2011 A model-based approach to determine the long-term effects of multiple interacting stressors on coral reefs. Ecol. Appl. 21, 2722–2733. (doi:10.1890/10-2195.1)

43. Blackwood JC, Hastings A, Munday PJ. 2012 The effect of fishing on hysteresis in Caribbean coral reefs. Theor. Ecol. 5, 105–114. (doi:10.1007/s12080-010-0102-0)

44. Fattahpour H, Zangeneh HRZ, Wang H. 2019 Dynamics of coral reef models in the presence of parrotfish. Natl. Resourc. Model. 32, e12202. (doi:10.1111/nrm.12202)

45. Carturan BS, Pither J, Marichal J-P, Bradshaw CIA, Parrott L. 2020 Combining agent-based, trait-based and demographic approaches to model coral-community dynamics. eLife 9, e55993. (doi:10.7554/eLife.55993)

46. Melbourne-Thomas J, Johnson CR, Aliño PM, Geronimo RC, Villanoy CL, Gurney GG. 2011 A multi-scale biophysical model to inform regional management of coral reefs in the western Philippines and South China Sea. Environ. Modell. Softw. 26, 66–82. (doi:10.1016/j.envsoft.2010.03.033)

47. Garin A, Tazir G. 2019 A topological ‘reading’ lesson: classification of MNIST using TDA. In 18th IEEE Int. Conf. on Machine Learning and Applications, Boca Raton, FL, USA, 16–19 December 2019, pp. 1551–1556. (doi:10.1109/ICMLA.2019.00026)

48. Carlsson G. 2020 Topological methods for data modelling. Nat. Rev. Phys. 2, 697–708. (doi:10.1038/s42254-020-00249-3)

49. Muthukrishnan R, Fong P. 2014 Multiple angiogenic stressors exert complex, interactive effects on a coral reef community. Coral Reefs 33, 911–921. (doi:10.1007/s00338-014-1199-1)

50. Davies N et al. 2016 Simulating social-ecological systems: the Island Digital Ecosystem Avatars (IDEA) consortium. GigaScience 5, 1–4. (doi:10.1186/s13742-016-0118-5)

51. Collin A et al. 2018 Very high resolution mapping of coral reef state using airborne bathymetric lidar surface-intensity and drone imagery. Int. J. Remote Sens. 39, 5676–5688. (doi:10.1080/01431161.2018.1500072)

52. Nardini JT, Stolz BJ, Flores KB, Harrington HA, Byrne HM. 2021 Topological data analysis distinguishes parameter regimes in the Anderson-Chaplain model of angiogenesis. PLoS Comput. Biol. 17, e1009694. (doi:10.1371/journal.pcbi.1009694)

53. Bhaskar D, Manhart A, Milzman J, Nardini JT, Storey KM, Topaz CM, Ziegelmeier L. 2019 Analyzing collective motion with machine learning and topology. Chaos 29, 123125. (doi:10.1063/1.5125493)

54. Thorne T, Kirk PDW, Harrington HA. 2022 Topological approximate Bayesian computation for parameter inference of an angiogenesis model. Bioinformatics 38, 2529–2535. (doi:10.1093/bioinformatics/btac118)

55. McDonald RA, Neuhäuser R, Robinson M, Larsen LG, Harrington HA, Bruna M. 2023 Zigzag persistence for coral reef resilience using a stochastic spatial model. Figshare. (doi:10.6084/m9.figshare.c.6793690)