



# EPA Public Access

Author manuscript

*Ecol Inform.* Author manuscript; available in PMC 2024 October 07.

About author manuscripts

Submit a manuscript

Published in final edited form as:

*Ecol Inform.* 2024 June 5; 82: . doi:10.1016/j.ecoinf.2024.102665.

## Exploring coral reef communities in Puerto Rico using Bayesian networks

John F. Carriger<sup>a,\*</sup>, William S. Fisher<sup>b</sup>

<sup>a</sup>U.S. Environmental Protection Agency, Office of Research and Development, Center for Environmental Solutions and Emergency Response, Cincinnati, OH 45268, USA

<sup>b</sup>U.S. Environmental Protection Agency, Office of Research and Development, Center for Environmental Measurement and Modeling, Gulf Breeze, FL 32561, USA

### Abstract

Most coral reef studies focus on scleractinian (stony) corals to indicate reef condition, but there are other prominent assemblages that play a role in ecosystem structure and function. In Puerto Rico these include fish, gorgonians, and sponges. The U.S. Environmental Protection Agency conducted unique surveys of coral reef communities across the southern coast of Puerto Rico that included simultaneous measurement of all four assemblages. Evaluating the results from a community perspective demands endpoints for all four assemblages, so patterns of community structure were explored by probabilistic clustering of measured variables with Bayesian networks. Most variables were found to have stronger associations within than between taxa, but unsupervised structure learning identified three cross-taxa relationships with potential ecological significance. Clusters for each assemblage were constructed using an expectation-maximization algorithm that created a factor node jointly characterizing the density, size, and diversity of individuals in each taxon. The clusters were characterized by the measured variables, and relationships to variables for other taxa were examined, such as stony coral clusters with fish variables. Each of the factor nodes were then used to create a set of meta-factor clusters that further summarized the aggregate monitoring variables for the four taxa. Once identified, taxon-specific and meta-clusters represent patterns of community structure that can be examined on a regional or site-specific basis to better understand risk assessment, risk management and delivery of ecosystem services.

### Keywords

Coral reefs; Community ecology; Bayesian networks; Cluster analysis

---

This is an open access article under the CC BY-NC-ND license (<https://creativecommons.org/licenses/by-nc-nd/4.0/>).

\*Corresponding author. carriger.john@epa.gov (J.F. Carriger).

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

CRedit authorship contribution statement

**John F. Carriger:** Conceptualization, Formal analysis, Investigation, Methodology, Project administration, Validation, Visualization, Writing – original draft, Writing – review & editing. **William S. Fisher:** Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Project administration, Resources, Supervision, Visualization, Writing – original draft, Writing – review & editing.

## 1. Introduction

Coral reefs occur globally in tropical waters and are formed by colonies of scleractinian (stony) corals that secrete calcium carbonate skeletons as they grow (Chave et al., 1972; Sorokin, 1995). This structure provides habitat for a variety of organisms, including fish and invertebrates, that combine to form the coral reef ecosystem (Maragos et al., 1996; Reaka-Kudla, 2005; Roberts et al., 2002). The presence of scleractinian (stony) coral colonies affects currents, waves and light penetration in the water column and creates niches and micro-niche habitats for diverse organisms and community interactions that effect primary and secondary production, calcium carbonate production and erosion, and metabolic exchange (Dennison and Barnes, 1988; Bruno and Bertness, 2001; Bellwood et al., 2019; Brandl et al., 2019; LaRue et al., 2023a, b). Most coral reef studies logically focus on the foundational assemblage, i.e., stony corals, but other reef components influence ecosystem structure and function. Reef ecosystems along the southern coast of Puerto Rico have four prominent biological assemblages—stony corals, soft corals (gorgonians), sponges, and fish. Combined, these assemblages and their interaction play roles in primary and secondary production, metabolic exchange, carbon sequestration, wave attenuation and sand production, among others (Mumby et al., 2008; Yee et al., 2014b). Two surveys were performed in southern Puerto Rico, one in 2010 and one in 2011, to characterize the reef community (Fisher et al., 2019). The surveys were novel in several respects: They measured density, size and morphology of individual benthic colonies rather than traditional two-dimensional cover estimates, sampled the entire southern coast of Puerto Rico rather than a specific reef location, and for the first time collected data for all four assemblages simultaneously. These factors generated a unique dataset of reef community structure in Puerto Rico and an opportunity to explore patterns and potential relationships among the assemblages.

Patterns in reef structure were generated using the quantitative relationships among measured variables, including diversity, number, size and shape of individuals from all four assemblages. Such information can lead to improved understanding of risk assessment (Alvarez-Filip et al., 2011; Carriger et al., 2021; Oliver et al., 2011) and risk management (Bradley et al., 2020; Santavy et al., 2022a, 2022b) during a prolonged period of stony coral decline (Carpenter et al., 2008; Dubinsky and Stambler, 1996; Hoegh-Guldberg et al., 2017; Kleypas and Yates, 2009; Knowlton and Jackson, 2008; Mora, 2008). It may also lead to a better quantification of ecosystem services provided by the different reef components (Spurgeon, 1992, Moberg and Folke, 1999, Beaumont et al., 2008, Pendleton, 2008, Principe et al., 2012, Yee et al., 2014a, 2014b. Principe and Fisher, 2018, Woodhead et al., 2019, Kelutur et al., 2021).

Ecosystem services are provided by all four predominant biological assemblages in Puerto Rico: Corals and gorgonians provide habitat; fish provide food; corals, fish and gorgonians provide tourism opportunities; corals and gorgonians provide coastal storm protection; and gorgonians and sponges provide novel marine biochemicals. Number, size and diversity of individuals are three basic measurements that can be used to evaluate delivery of these services for each taxon. For example, size and density can indicate potential for habitat

and for coastal storm protection, biomass for food production, and density and diversity for tourism opportunities and marine product discovery. While these measures can estimate the contribution of each taxon to a service, it is unknown whether the presence and condition of other taxa at a reef will influence (amplify or attenuate) that contribution. Some potential relationships are foreseeable. For example, stony corals, gorgonians and sponges may compete to colonize benthic surfaces (Dahl, 1973, Cruz et al., 2016, Ladd et al., 2019). Yet all three can also provide surface area above the sea floor that can be used by fish and invertebrates as habitat, providing shade, predator avoidance and opportunity for predation (Lirman, 1999, Syms and Jones, 2000, Darling et al., 2017).

A colony-based (demographic) survey approach is being used by several stony coral assessment programs (Kramer, 2003, Fisher et al., 2007, Fisher et al., 2014, NOAA, 2014, FRRP, 2022). This approach records the species, height, maximum diameter, and proportion of live tissue on each colony in a transect. Using this demographic approach, scientists from the Environmental Protection Agency (EPA) performed two assessment surveys on stony corals at multiple sites in Puerto Rico (Fisher et al., 2019). The number, morphology and size of sponge and gorgonian colonies and the number, species, and size of fish were documented simultaneously at each transect. The multiple variables measured in these surveys provide an opportunity to examine potential relationships among the four co-located taxa and generate hypotheses for patterns in reef communities, reef integrity and the delivery of ecosystem services.

Translating biological status from multiple monitoring variables requires advanced statistical approaches. Classical approaches rely on linear relationships between variables, but these are less reliable for environmental datasets where non-linear relationships may exist. A Bayesian network approach that uses a nonparametric graphical representation of a joint probability distribution for the variables in a model (Pearl, 1988) was employed for this study. This approach allows for clustering of multiple variables and drawing inferences from resulting joint probabilities. The network consists of nodes or graphical random variables connected by arcs (directed arrows) that indicate a quantitative relationship between variables, which are characterized as a receiving node (child) or an originating node or group of nodes (parent (s)). The relationships between directly connected nodes are often developed in conditional probability tables that contain the frequencies of parent-child relationships. Inference is gained by incorporating hard or soft evidence on one or more nodes, thereby updating the joint probabilities throughout the model using calculus and the conditional independencies in the network. Inferences about the probability of one node being in a particular state given a set of evidence on other nodes generates opportunity for systematic exploration, which can be made in multiple directions. Sensitivity among nodes that are not directly connected can be examined through the inferential process, which is one reason that Bayesian networks are preferred for reasoning with complex problems (Conrady and Jouffe, 2015). Other key advantages are the transparency of the model structure and relationships which invites great capabilities for using data with expert input. As an example, Ban et al. (2014) chose a Bayesian modeling platform to incorporate expert knowledge where data were lacking for multiple stressors on corals in the Great Barrier Reef.

Recent quantitative studies with coral reefs have emphasized varying machine learning approaches to assess sites and regions that lack sufficient monitoring data. Numerous studies have used machine learning for image analysis of coral reef components (Rubbens et al., 2023). These have been primarily applied to coral species identification (Ganesan and Santhanam, 2022; Marre et al., 2020; Mills et al., 2023; Sun et al., 2022; Villon et al., 2021), but also include reef community status (Gonzalez-Rivero et al., 2020; Marre et al., 2020; Schürholz and Chennu, 2023) and habitat mapping in support of coastal management (e.g., Barve et al., 2023; Do and Tran, 2023; Da Silveira et al., 2021). Some studies have shown that multiple machine-learning networks may be more successful at reef image identification (Burns et al., 2022; Lumini et al., 2020; Sun et al., 2022). Rubbens et al. (2023) have reviewed applications of machine learning to marine ecology, including coral reefs.

Cluster analysis has been used extensively to classify coral reefs for a variety of purposes. These include, but are not limited to, assessing wave energy in relation to benthic functional groups (Ford et al., 2021), predicting and monitoring coral bleaching events (Boonnam et al., 2022; Ford et al., 2024), reef acoustics (Ozanich et al., 2021), larval connectivity (Burt et al., 2024), habitat types (Barve et al., 2023) and genetic species differentiation (Meziere et al., 2024). Donovan et al. (2018) clustered reef habitat and fish data for Hawaii using expectation-maximization clustering with Bayesian model selection features. Their work used visual survey data to identify distinct habitat regimes that integrated fish with benthic reef features.

Machine learning and clustering approaches are used to characterize community-level features that are not readily evident from single component analyses. Community characteristics and interactions play a role not only in reef integrity but also in the provision of ecosystem services. Consequently, any means to capture community-level patterns can generate a deeper understanding and a greater potential for successful management. To this end, Bayesian networks were used for clustering community assemblages based on multi-dimensional measurements of physical structures of coral reef and characteristics of fish assemblages. A new factor node was created from the clustering and employed for examining the relationship with measured coral reef characteristics for interpreting the cluster states. Individual factor nodes were created for each taxa (sponges, gorgonians, stony corals, fishes) and then a meta-factor was constructed from clustering the taxa-specific factor nodes to summarize overall characteristics of the reef. The relationships between the factor nodes and measured characteristics for taxa used in their construction, and other species taxa, identified and provided new information on cross-taxa relationships between different reef types. There are both challenges and opportunities for using Bayesian networks to cluster biological monitoring data and generate inferences on ecological integrity. In this study, we first applied machine learned Bayesian networks and probabilistic clustering analysis to explore monitoring data for coral reef ecosystems. By examining the structure of the networks and the quantification of the parameters in this process, we elucidated key associations among coral reef indicators that may be used to hypothesize causal relationships. This phase involved unsupervised examination of how the measured variables related to one another and highlighted features related to cross taxa relationships. In the second phase, clustering was used to group the measured coral reef attributes. Clustering reduced multiple reef measurements into fewer dimensions, thereby facilitating

the interpretation of distinct community patterns across sites. Relationships identified in the exploratory phase were advanced by comparing clusters across taxa. The final product of this phase was a hierarchical Bayesian network that included all clusters and measured variables. Including meta-clusters that grouped the taxa-specific clusters. to represent community composition across all four reef components.

## 2. Methods

### 2.1. Data set

Two coral reef assessment surveys were completed along the southern coast of Puerto Rico (Fisher et al., 2019). The first survey was completed on Nov 29-Dec 13, 2010 (PR10). This survey used a targeted design that included 74 stations at depths of 1–12 m. The second survey was completed on Nov 27-Dec 12, 2011 (PR11). This survey used a random survey design that included 64 stations at depths of 1–12 m. The combined dataset totaled 138 sites at 1–12 m depth. However, some stations extended to greater depths (Fig. S11). For both surveys, four biological assemblages (fish, stony corals, gorgonian corals, and sponges) were documented simultaneously along a single transect line using procedures outlined in Fisher (2007) and Santavy et al. (2012). For fish, the number, species, and size class were recorded along a 4 m × 25 m (100 m<sup>2</sup>) transect area. Species and dimensions (height and maximum diameter) for each stony coral colony were recorded within a 1 m × 15 m (15 m<sup>2</sup>) transect area for PR10 and a 1 m × 25 m (25 m<sup>2</sup>) transect area for PR11. Gorgonians and sponges were recorded in a 1m<sup>2</sup> quadrat placed five times at regular intervals (0-, 5-, 10-, 15- and 20-m marks) along the transect line. Both gorgonian and sponges were characterized by morphology (nine morphologies for gorgonians and eight for sponges; Santavy et al., 2012) and dimensions (height and maximum diameter). Data for each station were normalized to m<sup>2</sup>. Field data from the surveys were amended to exclude hydrocorals and stony coral colonies with no live tissue. Field and amended datasets are available at EPA's Environmental Dataset Gateway (U.S. EPA, 2024).

### 2.2. Variable descriptions

Fish biomass estimates were calculated from 5-cm size classes according to the formula  $Biomass = \alpha L \beta$  where L is fish length (cm, midpoint of size class) and  $\alpha$  and  $\beta$  are species-specific length-weight relationships found in FishBase (Froese and Pauly, 2007). Variables for fish populations included fish density (FDn, n m<sup>-2</sup>), taxa richness (F-TR, families m<sup>-2</sup>), fish biomass (F-Bm, g m<sup>-2</sup>) and average biomass per fish (F-Bmf).

Size characteristics for stony corals were estimated using a hemisphere as a geometric surrogate (Fisher et al., 2014). Colony footprint (C-Fp) was calculated as  $\pi r^2$  where radius  $r = \frac{1}{2}$  maximum diameter. Colony surface area (C-SA) was calculated as  $M'(r')^2$  where M = 1, 2, 6 or 8 for flat, hemispheric, lobed/globular or branched species (Fisher et al., 2014). To better account for colony height'  $r' = (\text{radius} + \text{height})/2$  was substituted for colony radius (r) in C-SA calculations. Surface Index (Dahl, 1973) was calculated as C-SA/C-Fp and volume occupied (C-VO) as  $\frac{2}{3}\pi r^2 h$ . Stony coral variables calculated for each station included density (C-Dn, n m<sup>-2</sup>) taxa richness (C-TR, species m<sup>-2</sup>), means and sums of

height (C-Hmn, cm; C-Hsum, cm m<sup>-2</sup>), and sums of footprint (C-Fp, cm<sup>2</sup> m<sup>-2</sup>), surface area (C-SA, cm<sup>2</sup> m<sup>-2</sup>), surface index (C-SI), and volume occupied (C-VO, cm<sup>3</sup> m<sup>-2</sup>).

Surface area and volume occupied by gorgonians were estimated using either a cylinder or a cube as a geometric surrogate. Footprint, surface area and volume were calculated as a cylinder ( $Fp = \pi r^2$ ;  $SA = 2\pi rh + \pi r^2$ ;  $Vol = \pi r^2 h$ ). However, for planar sea fans and planar sea rods, footprint, surface area and volume were calculated as a cube ( $Fp = d*w$ ;  $SA = 2(h*d) + 2(h*w) + 2(w*d)$ ;  $Vol = h*d*w$ ) where  $d$  is maximum diameter and width ( $w$ ) was assigned as 1 cm. Gorgonian variables included density (G-Dn, n/m<sup>2</sup>), morphological richness (G-MR, morphologies/m<sup>2</sup>), and means and sums of height (G-Hmn, cm; G-Hsum, cm m<sup>-2</sup>), footprint (G-Fp, cm<sup>2</sup> m<sup>-2</sup>), surface area (G-SA, cm<sup>2</sup> m<sup>-2</sup>), surface index (G-SI = G-SA/G-Fp) and volume occupied (G-VO, cm<sup>3</sup> m<sup>-2</sup>).

Footprint, surface area and volume occupied by the eight sponge morphologies were estimated using a cylinder as a geometric surrogate ( $Fp = \pi r^2$ ;  $SA = 2\pi rh + \pi r^2$ ;  $VO = \pi r^2 h$ ), where radius ( $r$ ) = 1/2 maximum diameter. Surface index was calculated as SA/Fp. Sponge variables included density (S-Dn, n m<sup>-2</sup>), morphological richness (S-MR, morphologies m<sup>-2</sup>) and means and sums of height (S-Hmn, cm; S-Hsum, cm m<sup>-2</sup>), and sums of footprint (S-Fp, cm<sup>2</sup> m<sup>-2</sup>), surface area (S-SA, cm<sup>2</sup> m<sup>-2</sup>), surface index (S-SI), and volume occupied (S-VO, cm<sup>3</sup> m<sup>-2</sup>).

### 2.3. Bayesian network analysis

All Bayesian network analysis was conducted with BayesiaLab 10.2 (Bayesia S.A.S., 2022) using approaches described in the BayesiaLab help function and Conrady and Jouffe (2015). Coral reef monitoring variables were discretized prior to analysis. The R2-GenOpt procedure in BayesiaLab for discretization was used to reconstruct the existing patterns in the data. The algorithm uses genetic optimization to maximize the R<sup>2</sup> between the original raw data distribution and the density created by the discretization. Three intervals were selected following the recommended discretization number in BayesiaLab as calculated from the size of the data set. Discretization intervals are displayed in Table S1. Quantification of the parameters was developed through maximum likelihood estimation of the parent/child state occurrences in the dataset. This uses the counts of the data for establishing a probability for each child state given a combination of parent states.

**2.3.1. Exploratory analysis**—Distance mapping with Pearson correlation and mutual information were used for the unconnected nodes to examine how the manifest (observed) variables clustered based on the strength of their relationships. This also helped to compare relationships directly in the data with the model structure. The distance mapping procedure uses a measure of sensitivity to determine the proximity of a collection of variables across the layout of the BayesiaLab window. Distance mapping for Pearson correlation is based on the coefficient value with two positively correlated nodes being grouped closer together and two negatively correlated nodes being grouped at opposite ends of the workspace. Mutual information quantifies how much uncertainty in a node is reduced when the value of a second node is known. Distance mapping with mutual information shows nodes having higher mutual information clustering closer together.

Maximum weight spanning tree (MWST) structure learning was used to gain additional insights. The MWST is considered a tree structure because it is constrained to only one parent node per child node. The MWST starts off by examining the strength of the relationships between the nodes and connects them by prioritizing the strongest relationships. The direction of the arcs is automatically chosen to minimize structural complexity (Bayesia S.A.S., 2010). The MWST networks were built using a score-based determination from the minimum description length (MDL) (Conrady and Jouffe, 2015). The MDL contains two components—one for the representation of the data and one for the complexity of the network. The complexity term is given a weight, known as a structural coefficient (SC). When the SC is 1, the terms are balanced equally between complexity and data representation. When the SC is  $>1$ , less complexity is favored and when the SC is  $<1$ , greater complexity is favored over the balanced network. A network with a balanced score (SC = 1) between complexity and data representation was first constructed. Then, the SC was reduced to increase complexity until all nodes were linked in a single network. The final structure was attained when reductions in the SC (in 0.05 intervals) first permitted a completely connected network.

Node importance was tested in the resulting structure from the connected MWST network. In this process, arc force was used to gain knowledge of the sensitivity of the relationships in the model. Arc force provides a Kullback-Leibler divergence measure based on the information brought to a joint probability distribution when the arc is included or removed (Conrady and Jouffe, 2015). The Kullback-Leibler divergence is a standard statistical tool in information theory that measures the proximity of two distributions. In this case, the network with the arc in question was compared to a hypothetical network that is in every way the same but missing the arc. This isolates and measures the contribution of the arc in the joint probability represented by the Bayesian network. The nodes that were most strongly connected to other nodes (node centrality) were found through the node force values which relies on the total arc force (sums of the incoming and outgoing arc forces) to examine the overall strength of the direct connections (Conrady and Jouffe, 2015). Higher node force values equate to higher centrality of nodes in the network. Pearson correlation analysis between each of the measured variables was also used to examine the linear strength and direction of the positive and negative relationships.

**2.3.2. Clustering analysis**—Stations were grouped into clusters based on their similarities (Conrady and Jouffe, 2014). The observed characteristics (manifest variables) of sponges, gorgonians, stony corals, and fishes at sites were used to group the sampling sites. This generated a latent (unobserved) node for each taxon that is distinct from the manifest observations. The procedure is used to construct new variables to summarize the taxa variables comparable to the extraction of latent or composite variables in a traditional factor analysis or principal component analysis (PCA). In this case, a latent factor node is created from the variables measured at the reefs that summarizes the associations among the sites. The algorithm utilizes the MDL score in a naïve Bayesian network to identify the latent variable with multiple states to represent clustering among the data that is connected to the measured variables used in its construction. A naïve Bayesian network contains a center node (the parent) with predictor nodes as children. In this case, the center node is

the factor node and the nodes for the observed reef variables are the predictors. The latent factor node was constructed through an expectation-maximization process that targeted the similarities in the monitoring sites for each of the candidate cluster groupings. Sherif et al. (2015) describe the expectation-maximization process as an approach that iterates back and forth between expectation and maximization steps. It starts with a ‘guess’ (expectation), then checks if this guess can be improved (maximization) and then iteratively re-guesses with the improved values. In this case, the algorithm initially starts with random distributions over the cluster and reef states. In the first expectation step, the clusters are imputed based on these starting distributions and the data. The maximization step uses the imputations from this step to realign the conditional probabilities in the network from the data. This process is repeated until convergence (no change) or a stopping point is reached.

The MWST-learned network was used for the final clustering of each group of variables. The connections between nodes for different taxa variables were first removed so that only within-taxon connections remained (Fig. S2). The number of factors requested for variable clustering was chosen based on the number of taxa. Multiple clustering was used to construct a new latent factor variable for each taxon, generating multiple cluster states within each latent factor variable. These four latent variables were then themselves clustered to create a higher-level latent factor variable, a meta-cluster, that summarized the cluster states across taxa. This meta-clustering provides a hierarchical clustering network that summarizes information on the coral reef community from the measured variables across taxa. This network also includes all factor nodes and observable nodes from monitoring data so it provides an overall network for the community structure based on the observed reef characteristics. The factor nodes can then be used to perform inferences in this network like factor scores from PCA may be used in a data case file (Conrady and Jouffe, 2014).

For the data clustering procedures, several constraints were applied. A minimum purity of 70% was set for accepting a taxon-specific cluster but this was increased to 85% for accepting a meta-cluster. Purity measures capture the highest posterior probability of a latent variable cluster state for the data on each of the coral reef monitoring nodes from a site. Thus a 100% pure site would be contained completely in one cluster but a lower purity would be partially contained in other clusters from the posterior probabilities. Purity measures are computed as an average for each cluster and associated site data and as an overall average across all site data and associated cluster states of a latent factor variable. The supporting material contains these purity measures and other statistics related to the clusters and resulting network (Table S4).

The cluster analysis was constrained to a maximum of five states with a minimum of two, but low purity measures can exclude some candidate clusters. The optimal number of cluster states and the data contained in each was identified using an automated expectation-maximization process, described above, based on a random walk with 300 steps for minimizing the MDL score. The final ten steps were used to check for stability with the number of clusters chosen for each taxon. A stable outcome would have the same number of clusters for all ten final steps. The contingency table fit (CTF), a network performance indicator representing the quality of the fit to the data, provides a normalized measure from 0 to 100% with 0% (poorest fit) having the data representation skill of an unconnected



network and 100% having the joint probability representation of the data of a completely connected network (best possible fit) (Conrady and Jouffe, 2014; Gerassis et al., 2019). The CTF was extracted for the individual taxa factor networks as well as the meta-factor network.

The patterns in the data were identified through the resulting naïve Bayesian network that was constructed for each community component (taxon) and the hierarchical factor node. The hierarchical network integrated all nodes together in one network with the latent factor node containing the meta-cluster states as the target variable, the factor nodes for each taxa as intermediaries, and the reef variables as manifest predictors. Once the data clustering was completed for each of the taxa and the hierarchical network, the contribution of each manifest predictor variable to the target factor node was identified, similar to comparing factor loadings for each variable in PCA (Conrady and Jouffe, 2014). Next, the relationships between the cluster states and the manifest variables' distributions were examined through a posterior means analysis (Bayesia S.A.S., 2022). The posterior means analysis examined the mean value of the manifest nodes for each of the cluster states of the latent factor. The procedure used for calculating a posterior mean for a measurement node given a particular cluster is further described in the supporting materials (CarrigerFisherSI\_PMA).

The coral and fish factor nodes were also examined outside of the observable nodes used in their creation. The coral factor node was connected separately to the fish manifest nodes and to the gorgonian manifest nodes in naïve Bayesian networks. The fish factor was also connected to all observable variables not used in its creation (i.e., for sponge, gorgonians, corals) in a naïve Bayesian network. The clustering process had assigned all the identified clusters to each site in the data set, thus allowing a new Bayesian network to be created that contains variables in the data not used for creating a cluster. A posterior means analysis was used for examining the relationship between the fish or coral clusters and the observable variables.

### 3. Results

#### 3.1. Exploratory analysis

Distance mapping provided a clear picture of the variables that are most closely related (Fig. 1). Both mutual information and Pearson correlation showed that within-taxon nodes grouped with one another. Pearson correlation found the strongest positive linear relationships between sponge and gorgonian variables, shown by the proximity of their nodes (Fig. 1a). Coral nodes were most negatively related to sponge nodes and the fish nodes were most negatively related to the gorgonian nodes. Mutual information found similar groupings to the Pearson correlation despite being a nonlinear measure (Fig. 1b). A subset of the coral variables (C-VO, C-Hmn, C-SA and C-Fp) was more centrally located in the mutual information map indicating stronger ties for those variables to other community components.

The MWST algorithm with SC set at 1 identified two connected structures (Fig. 2a). One structure contains the sponge and gorgonian nodes and the second contains the fish and coral nodes. For the former, S-MR is connected to G-SA and for the latter, F-DN is

connected to C-SI. These connections are reflected in the distance mapping with mutual information relationships where both node relationships are among the more closely spaced inter-variable relationships in the graph. The S-MR and G-SA relationship is also visible for the Pearson correlation distance map as being among the strongest inter-variable relationships indicating a more positive linear relationship between the two nodes. The SC value that provided a fully connected network was 0.80 (Fig. 2b). The new connection that formed with SC = 0.80 was between G-SI and C-Hmn and was the only negative Pearson correlation value in the MWST graphs (Table S2, Fig. S1). The Pearson correlation distance map indicated that these are two of the more negatively positioned nodes between the gorgonian and coral node groups. The flow among variables for corals and gorgonians seemed to exhibit similar patterns with the connections between Hsum, Dn, SI, MR and TR variables. Some of these patterns are also seen in the connections for the sponges.

The fully connected network was examined for node centrality (nodes with the most and strongest connections, Conrady and Jouffe, 2015). The node force measure, based on Kullback-Leibler divergence, provided a visual interpretation of node forces in the network (Fig. 3). Higher node force values are used to identify which nodes are central in the network. Numeric values for node force are provided in Table S3. The sum of heights for sponges, corals and gorgonians were key variables for those components, and surface area for sponges and gorgonians also had relatively high node force values. The fish variables, based on the relative magnitude of their node force values, appeared to play a smaller role overall in the joint probability of the network.

### 3.2. Clustering of reef variables

Latent factor nodes were developed from the measured variables. The contribution of each measured variable to its latent factor was identified for each taxon (Fig. 4). For sponges and gorgonians, colony footprint nodes (S-Fp and G-Fp) were the largest contributors to the latent factor and surface area was the second largest (S-SA and G-SA). For corals, mean height and colony footprint made the greatest contributions to the latent factor. For fish, the greatest contributions were from the two biomass variables—overall biomass and biomass per fish—and these provided >90% of the contribution to the latent factor. Five clusters were identified for each taxa and the CTF values were all equal to or above 70% for the individual taxa networks (Table S4).

Grouping of multiple states within the latent factors created five clusters with unique characteristics for each of the taxa (Tables 1–4, Figs. S3-S6). A tutorial on the calculations for posterior means and their interpretation is provided in the supporting materials (CarrigerFisherSI\_PMA). For sponges, the third cluster (SC3) had the highest posterior mean values (Table 1) for all measured variables, but only 4.3% of the stations were in this cluster. The fourth cluster (SC4) had the lowest posterior mean values and constituted 24.6% of the stations. Among the five fish clusters, FC2 and FC3 split the higher posterior mean values with FC2 having the higher fish biomass variables and FC3 having higher values for fish density and taxa richness (Table 2). Combined, these two clusters constituted only 5.1% of the stations. The lowest posterior mean values for fish were predominantly found in FC4 and constituted 50% of the stations. Most of the highest posterior mean

values for gorgonians were in clusters GC4 and GC5 (Table 3), comprising 19.6% and 7.9% of stations, respectively. The lowest posterior means for gorgonians were in Cluster GC2, comprising 14.5% of stations. Coral had the highest posterior mean values in CC1 and CC5, comprising 12.3% of stations and the lowest values in CC4 representing over 50% of stations (Table 4). Cluster 1 for the corals (CC1) had the lowest value for C-TR and CC3 had the lowest posterior mean value for C-Hmn. The latter also had relatively high posterior mean values for taxa richness and density. The highest percentage of stations were associated with lower posterior mean values for sponges, fish and corals. For gorgonians, ~28% of stations were associated with the highest posterior mean values (GC4 and GC5).

### 3.3. Cluster comparisons

Detaching the fish factor node from the fish variables and attaching it to the manifest variables of other taxa (Fig. 5) allowed comparison of each fish cluster (shown in Table 2) with other measured reef variables (Table 5, Fig. S7). Fish cluster 1 (FC1), which had lower posterior mean values for fish biomass, was associated with the highest values for C-SA and C-Fp. Fish cluster 2 (FC2) had the highest posterior mean values for fish biomass and was associated with the highest posterior mean values for C-Dn, S-FP, and G-Hmn (Table 5). Fish cluster 3 (FC3) had the highest posterior mean values for fish density and fish taxa richness, and was associated with highest C-Hmn, C-VO, C-Hsum and C-SI posterior means and the second highest C-Dn, C-Fp, S-Fp, and C-SA posterior mean values. It was also associated with lower posterior mean values for many of the gorgonian and sponge variables and C-TR. Fish cluster 4 (FC4), which had relatively low posterior mean values for all four measured fish characteristics, was associated with the lowest posterior mean values for C-SI, C-Dn and C-Hsum but with higher posterior mean values for C-TR and most other non-coral variables. Fish cluster 5 (FC5) exhibited moderate posterior mean values and was associated with the highest sponge values for S-Dn and S-Hsum; and the lowest for S-VO, S-Fp, and C-TR.

The coral factor node (CF) was also attached separately to the fish and gorgonian manifest nodes for analysis (Tables 6–7, Figs. S8-S9). Coral cluster 1 (CC1), which had higher posterior mean values for colony size (shown in Table 4), was associated with the highest F-Dn and F-TR while CC3, higher for coral taxa richness and density, was associated with the highest fish biomass means (Table 6). Coral cluster 4 (CC4) had lower mean values for all coral variables, and was associated with the lowest F-Dn and F-TR mean values. Coral cluster 5 (CC5) had the highest C-Hsum, C-Dn, and C-TR posterior mean values and was associated with the lowest fish biomass values. In relation to gorgonians, CC1 was associated with the lowest gorgonian posterior mean values and CC3, CC4, and CC5 with most of the higher gorgonian values (Table 7).

### 3.4. Hierarchical clustering

The hierarchical clustering initially identified four clusters after 300 steps. However, the algorithm did not settle on a consistent number of clusters and 30,000 steps were used to stabilize the algorithm. In both cases, four clusters were identified, but the 30,000 steps clustering network was applied for analysis. The CTF for the meta-factor network was lower

than for the individual taxa at approximately 58% (Table S4) reflecting that the meta-factor network has high complexity and includes all four taxa factor and manifest nodes.

After 30,000 steps, hierarchical clustering identified a meta-factor variable that summarized factors for each community component (Fig. 6). The posterior mean analysis identified the most likely values for measured reef nodes for each of the meta-clusters (Table 8, Fig. S10). Meta-cluster 1 (MC1) exhibited the highest gorgonian and sponge posterior mean values and lower coral and fish values. The lowest coral posterior mean values and a majority of the lowest sponge values were seen in MC2. Gorgonians and fishes had moderate to low posterior mean values in MC2. Both MC1 and MC2 also contained most of the deeper sites sampled (>8 m) (Fig. S11). The highest fish biomass measurements and coral density and taxa richness and the majority of the lowest gorgonian values were in MC3. The F-Dn and F-TR posterior mean values were lowest in MC3 and the rest of the posterior mean values for coral nodes were relatively moderate. Sponges and gorgonian posterior mean values were mixed between relatively low, moderate, and high values in MC3. The highest posterior mean values for coral variables, except density and taxa richness, along with the highest fish density and taxa richness mean values, were in MC4. The latter also had relatively low to moderate posterior mean values for gorgonian and sponge nodes. The attributes of these four meta-clusters are illustrated by photographs taken at representative sites (Fig. 7): MC1 stations (44.9% of stations sampled) showed higher density and diversity of gorgonians and sponges and low variable values for fish and coral. Meta-cluster 2 (MC2) stations (35.5%) were largely devoid of all benthic organisms but exhibited moderate fish values. Meta-cluster 3 (MC3) stations, which comprised only 8.0% of the stations, had higher coral density and taxa richness coincident with higher fish biomass. Meta-cluster 4 (MC4, 11.6% of stations) had higher coral values for physical variables, moderate coral taxa richness, and higher fish density and taxa richness.

#### 4. Discussion

Linear and non-linear associations of four predominant taxa forming coral reef communities along the southern coast of Puerto Rico were examined using a novel Bayesian network application to distinguish community patterns from density, diversity, and size relationships. An exploratory analysis with MWST structures found the strongest correlations were within taxa, likely reflecting the interrelatedness of dimension variables (height, footprint, surface area, and volume) for each taxon. However, an MWST structure learning process also identified positive cross-taxon correlations between sponge and gorgonian variables and between fish and coral variables. Additionally, a negative Pearson correlation was found between coral and gorgonian nodes, but only after allowing greater complexity in the model to generate a completely connected network (Fig. 2b).

Examination of the connecting nodes can provide insight into the resulting interrelationships. For example, the node for F-Dn connected to the node for C-SI (Fig. 3), which is an estimate of the surface area of a colony relative to its footprint. This positive connection may reflect the provision of coral structural habitat for fish. The highest posterior mean values for both nodes were found in the same meta-cluster (MC3, Table 8), demonstrating how the meta-clusters encompass locations with the underlying

cross-taxa relationships. Positive connecting nodes between sponge and gorgonian variables were S-MR and G-SA, the basis for which is uncertain but hierarchical clustering also identified a single meta-cluster (MC1) with the highest posterior mean values for both nodes. C-Hmn and G-SI were the connecting nodes for a negative relationship between coral and gorgonians, indicating smaller coral colony heights at locations with greater gorgonian surface index. This may reflect physical shading by gorgonians to block coral photosynthesis and growth. However, there were also indications of a negative relationship between C-Fp and G-Fp that could indicate competition for substrate. Assuming either or both of these are correct, the apparent ability of gorgonians to withstand climate change factors throughout the Caribbean (Edmunds et al., 2016; Lenz et al., 2015; Ruzicka et al., 2013) could magnify the relatively deleterious climate effects on stony corals and lead to a community shift toward gorgonians in Puerto Rico reefs. The highest posterior mean value for C-Hmn and the lowest posterior mean value for G-SI were both found in MC3; likewise, the second lowest posterior mean value for G-Fp and the highest for C-Fp were found in MC3. These and other inter-assemblage relationships may be critical to understand emerging risk management approaches based on multiple assemblages (Bradley et al., 2020; Fore et al., 2008; Fore et al., 2009; Santavy et al., 2022a, 2022b).

Clustering analysis, which generates groupings based on similarity of variables across the 138 sampling stations, was first applied to individual taxa to illustrate single-taxon population patterns. It is worth noting that the purity setting for this analysis was 70%, meaning that observed values in the predictor states had at least a 70% probability of being in a particular cluster on average. All clusters met this requirement with average purities ranging from 98.18 to 99.95% across the four taxa (Table S4). The overall average purity for the meta-factor's clusters was 94.55%, underscoring relatively strong distinctions among the clusters. Clusters with the highest posterior mean values for fish, corals and sponges represented the fewest stations. That is, stations with higher posterior means for density, size and diversity for these three taxa were relatively rare. Gorgonians, however, had a more balanced station distribution for clusters of high, medium, and low values. This underscores the broad distribution of gorgonians across Puerto Rico reefs and re-emphasizes the differential effect that climate change factors might exert on gorgonians and scleractinians (Edmunds et al., 2016; Johnson and Hallock, 2020). Examination of the clusters can reveal several transect characteristics. For example, Table 4 shows that transects with higher density (CC5) have smaller colonies; and that transects with large colonies (CC1) have the lowest taxa richness. This within-taxa information can be valuable for highlighting rare areas or habitats (low-occurrences), or areas providing greater ecosystem benefits (high posterior mean values).

Community patterns were characterized by generating a latent factor for each taxon to compare with variables of other taxa (Fig. 4). A focus on coral relationships is warranted because of the centrality of coral variables in mutual information distance mapping (Fig. 1). Among coral clusters, CC1 was higher for most coral variables (Table 4) and when the latent coral factor (CF) was applied, posterior mean values for fish density and taxa richness were also relatively high (Table 6) but relatively low for gorgonians in association with CC1 (Table 7). In contrast, posterior means for corals were relatively low in CC4, and when CF was applied fish variables were relatively low and gorgonian variables were

relatively high. These comparisons emphasize the positive and negative relationships seen at some sites of coral with fish and gorgonians, respectively, and identify the most influential variables in each. As another example, a comparison of CC1 (lower density, larger coral colonies) with CC5 (higher density, smaller coral colonies) could illustrate a fish habitat advantage for locations with a few larger coral colonies over those with several smaller colonies. Correspondingly, Fisher (2023) found that small colonies provide little fish habitat. Reinforcing findings from the MWST analysis noted above, the relation of lower coral variables (CC4) with taller gorgonians might reflect shading by gorgonians or competition with corals for substrate at those sites. In contrast, the negative relationship between fish and gorgonians is confounding given the potential for gorgonian structure to serve as fish habitat, much like seagrass structure (Jones et al., 2021). A possible explanation is that thick densities of gorgonians found at many sites obscured smaller fish from surveyor counts. Or reef fish may not use gorgonians for habitat or may even avoid gorgonians because of anti-predator mucus secretions or structural spicules (Brown and Bythell, 2005; Harvell and Fenical, 1989; Pawlik et al., 1987).

The clustering process was applied a second time to generate meta-clusters that identify community patterns across the sampling stations. The meta-clustering integrated both benthic and fish taxa clusters to define clusters in a manner that is “inclusive of their dynamics” (Donovan et al., 2018). This process simultaneously incorporated all variables from all the taxon-specific clusters into a single indicator of the aggregate community. Using a posterior means analysis, it was possible to demonstrate which manifest variables were most influential for defining community characteristics. For example, MC1 contained higher gorgonian and sponge values and lower fish density, taxa richness and biomass. Similarly, MC4 had higher coral posterior mean values and fish density and taxa richness but lower gorgonian and sponge values. Interestingly, the posterior mean values for fish variables were higher in MC2 (second highest for each variable), which contained relatively low posterior means for benthic organisms, than for MC1 with higher gorgonian and sponge values. The meta-clusters grouped key relationships across taxa and captured several key relationships in the inter-cluster posterior mean comparison. For example, CC3 had higher density and taxa richness of corals but also had higher fish biomass posterior means when applied to the fish predictors and these higher posterior means were grouped under MC3. Likewise, the higher posterior mean values for the majority of the coral variables were found in CC1 which also was associated with higher fish density and taxa richness when connected to the fish predictors. These inter-relationships were also observed for the posterior mean values of MC4. The meta-cluster groupings contain all the measured characteristics of the reef systems and can differentiate sites based on measurements from multiple endpoints representing multiple taxa.

Hierarchical clustering of higher and lower assemblage characteristics can be useful as endpoints in future environmental assessments, not only as a community indicator but to reinforce the use of other assessment endpoints, such as presence of endangered species, reefscape attributes and single population biological variables that indicate provision of ecosystem services or vulnerability to stressors. Meta-clustering can be applied, with additional information, in ecological risk assessments for defining and characterizing communities at risk. Associations between reef characteristics and variables that target reef

structure and function could also be evaluated using both component (single assemblage) and meta-clusters. If the data are representative, the meta-clusters can be interpreted as different ecosystem states, relating reef structure to function across time and space. Greater understanding of these states and their ecosystem services can aid in risk assessment and management. Management effectiveness, as well, could be gauged by the loss or gain of a site from one meta-cluster to another after intervention. Additional temporal understanding of community transitional and stable states along with knowledge of the influence of anthropogenic and abiotic factors on coral reef communities would aid in the development and evaluation of management actions (Donovan et al., 2018). The meta-factor node can be directly adapted for Bayesian network risk models to determine external factors that influence the biotic community and nodes that represent useful interventions by management.

Some drawbacks exist in this approach, including the use of discretization on continuous variables potentially causing a loss of information. However, the method for discretization used here provides a balance between capturing the range and amount of data. The higher data requirements permit use of nonparametric analysis with more flexible assumptions than those found in other approaches. Nonetheless, analyses can be difficult with noisy environmental data, especially when the data sets are small. In addition, interpretation of the clusters contained in the latent factor variables for the community components is of utmost importance for their usefulness to environmental assessments. Although the current study found that the clusters provided clear delineation between higher and lower posterior means, the moderate values were difficult to interpret. This may be improved with better understanding of additional variables and characteristics of the site. Increasing the dimensions used in the assessment through additional variables can therefore be useful (Peterson and Evans, 2019). Due to the restrictions in MWST, the representation of data might not be as high as other learning methods (Kekolahti et al., 2015). However, the constraints (only one incoming connection per node) allow only the nodes with the strongest associations to be connected.

Clustering with Bayesian networks to understand ecological communities is a unique approach for coral reefs and other ecological communities, yet clusters can describe community patterns that are not intuitive nor otherwise discernible. The insights provided can be useful for ecological risk assessment and management or to interpret other data sets from a community perspective. Identification of habitat clusters for assemblages or species of concern provides a useful application of clustering with Bayesian networks where clusters can be built for one aspect of the data set and easily compared to another (e.g., fishes and coral). Future research can explore how the clusters identified with these monitoring data can be further compared to external stressors and in situ features that may be related to the status of reef communities.

## Appendix A. Supplementary data

Additional information available on the discretization thresholds, clustering results, statistical summary tables, and graphical output from the clustering analysis and the

resulting networks discussed in the article. Supplementary data to this article can be found online at <https://doi.org/10.1016/j.ecoinf.2024.102665>.

## Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

## Acknowledgments

The views expressed in this article are those of the authors and do not necessarily represent the views or the policies of the U.S. Environmental Protection Agency. Any mention of trade names, manufacturers or products does not imply an endorsement by the United States Government or the U.S. Environmental Protection Agency. EPA and its employees do not endorse any commercial products, services, or enterprises. This study was funded by the U.S. Environmental Protection Agency.

## Data availability

Data will be posted online at EPA website

## Glossary

### Fish

<b>F-Dn</b>	Density
<b>F-TR</b>	Taxa Richnes
<b>F-Bm</b>	Biomass
<b>F-Bmf</b>	Biomass/fish

### Coral

<b>C-Dn</b>	Density
<b>C-TR</b>	Taxa Richness
<b>C-Hsum</b>	Height Sum
<b>C-Hmn</b>	Height Mean
<b>C-Fp</b>	Footprint
<b>C-SA</b>	Surface Area
<b>C-SI</b>	Surface Index
<b>C-VO</b>	Volume Occupied

### Sponge

<b>S-Dn</b>	Density
<b>S-MR</b>	Morphological Richness
<b>S-Hsum</b>	Height Sum



<b>S-Hmn</b>	Height Mean
<b>S-Fp</b>	Footprint
<b>S-SA</b>	Surface Area
<b>S-SI</b>	Surface Index
<b>S-VO</b>	Volume Occupied
Gorgonian	
<b>G-Dn</b>	Density
<b>G-MR</b>	Morphological Richness
<b>G-Hsum</b>	Height Sum
<b>G-Hmn</b>	Height Mean
<b>G-Fp:</b>	Footprint
<b>G-SA</b>	Surface Area
<b>G-SI</b>	Surface Index
<b>G-VO</b>	Volume Occupied

## References

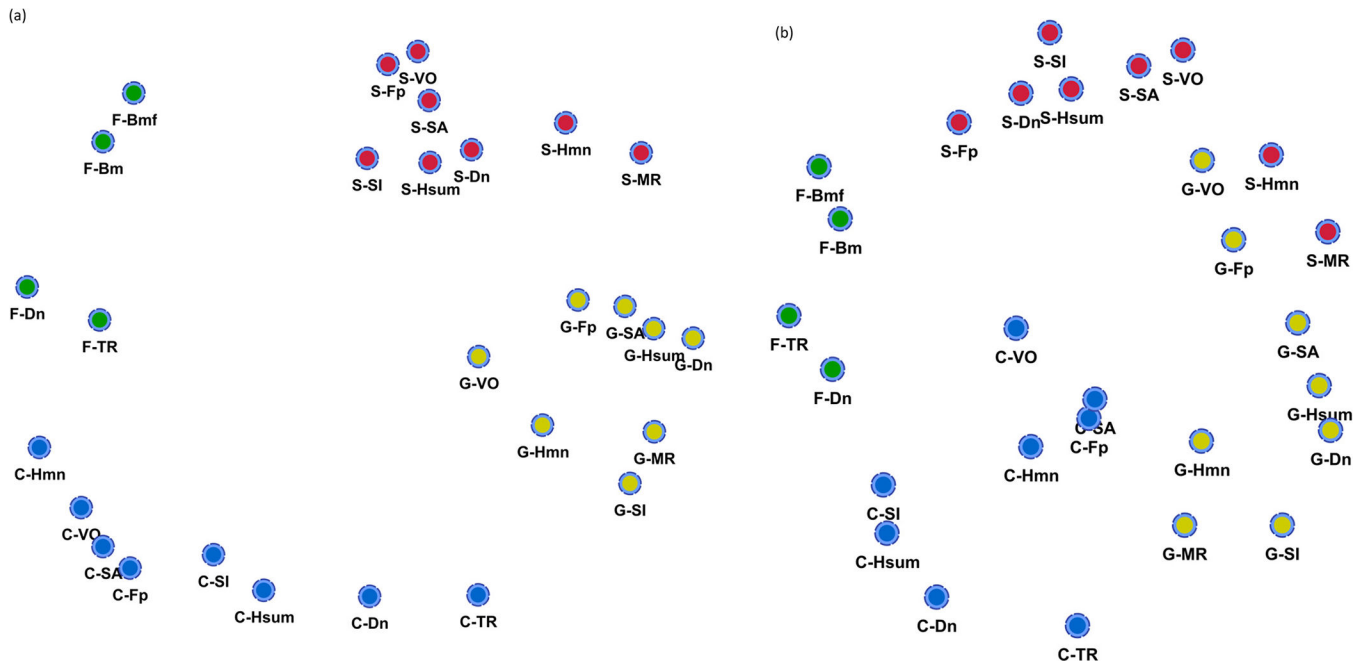
- Alvarez-Filip L, Gill JA, Dulvy NK, Perry AL, Watkinson AR, Côté IM, 2011. Drivers of region-wide declines in architectural complexity on Caribbean reefs. *Coral Reefs* 30, 1051–1060.
- Ban SS, Pressey RL, Graham NA, 2014. Assessing interactions of multiple stressors when data are limited: a Bayesian belief network applied to coral reefs. *Glob. Environ. Chang.* 27, 64–72.
- Barve S, Webster JM, Chandra R, 2023. Reef-insight: a framework for reef habitat mapping with clustering methods using remote sensing. *Information* 14 (7), 373.
- Bayesia SAS, 2010. BayesiaLab User Guide. Changé, France.
- Bayesia SAS, 2022. BayesiaLab Software Version 10.2 PE-L. Changé, France.
- Beaumont NJ, Austen MC, Mangi SC, Townsend M, 2008. Economic valuation for the conservation of marine biodiversity. *Mar. Pollut. Bull.* 56 (3), 386–396. [PubMed: 18191954]
- Bellwood DR, Streit RP, Brandl SJ, Tebbett SB, 2019. The meaning of the term “function” in ecology: a coral reef perspective. *Funct Ecol* 33, 948–961.
- Boonnam N, Udomchaipitak T, Puttinaovaratt S, Chaichana T, Boonjing V, Muangprathub J, 2022. Coral reef bleaching under climate change: prediction modeling and machine learning. *Sustainability* 14 (10), 6161.
- Bradley P, Jessup B, Pittman SJ, Jeffrey CFJ, Ault JS, Carrubba L, Lilyestrom C, Appeldoorn R, McField M, Schärer MT, Santavy DL, Wojtenko I, Smith T, Garcia G, Huertas E, Murry B, Walker BK, Ramos A, Gerritsen J, Jackson SK, 2020. Using reef fish as biocriteria to protect Caribbean coral reef ecosystems. *Mar. Pollut. Bull.* 159, 111287.
- Brandl SJ, Rasher DB, Cote IM, Casey JM, Darling E, Lefcheck JS, Duff JE, 2019. Coral reef ecosystem functioning: eight core processes and the role of biodiversity. *Front Ecol Environ* 17 (8), 445–454.
- Brown BE, Bythell JC, 2005. Perspectives on mucus secretion in coral reefs. *Mar. Ecol. Prog. Ser.* 296, 291–309.

- Bruno J, Bertness M, 2001. Habitat modification and facilitation in benthic marine communities. In: Bertness MD., Gaines SD. (Eds.), *Marine Community Ecology*. Sinauer Associates.
- Burns C, Bollard B, Narayanan A, 2022. Machine-learning for mapping and monitoring shallow coral reef habitats. *Remote Sens.* 14 (11), 2666.
- Burt AJ, Vogt-Vincent N, Johnson H, Sendell-Price A, Kelly S, Clegg SM, Head C, Bunbury N, Fleischer-Dogley F, Jeremie MM, Khan N, 2024. Integration of population genetics with oceanographic models reveals strong connectivity among coral reefs across Seychelles. *Sci. Rep.* 14 (1), 4936. [PubMed: 38472289]
- Carpenter KE, Abrar M, Aeby G, + 33 authors, 2008. One-third of reef-building corals face elevated extinction risk from climate change and local impacts. *Science* 321, 560–563. [PubMed: 18653892]
- Carriger JF, Yee SH, Fisher WS, 2021. Assessing coral reef condition indicators for local and global stressors using Bayesian networks. *Integr. Environ. Assess. Manag.* 17 (1), 165–187. [PubMed: 33200869]
- Chave KS, Smith SV, Roy K, 1972. Carbonate production by coral reefs. *Mar. Geol.* 12, 123–140.
- Conrady S, Jouffe L, 2014. Probabilistic Latent Factor Induction with Bayesialab: A Side-by-Side Comparison with Traditional Factor Analysis. Bayesia USA, Franklin, TN.
- Conrady S, Jouffe L, 2015. Bayesian Networks and Bayesialab: A Practical Introduction for Researchers. Bayesia USA, Franklin, TN.
- Cruz ICS, Meira VH, de Kikuchi RKP, Creed JC, 2016. The role of competition in the phase shift to dominance of the zoanthid *Palythoa cf. variabilis* on coral reefs. *Mar. Environ. Res.* 115, 28–35. [PubMed: 26849036]
- Da Silveira CBL, Strenzel GMR, Maida M, Gaspar ALB, Ferreira BP, 2021. Coral reef mapping with remote sensing and machine learning: a nurture and nature analysis in marine protected areas. *Remote Sens.* 13 (15), 2907.
- Dahl AL, 1973. Surface area in ecological analysis: quantification of benthic coral-reef algae. *Mar. Biol.* 23, 239–249.
- Darling ES, Graham NAJ, Januchowski-Hartley FA, Nash KL, Pratchett MS, Wilson SK, 2017. Relationships between structural complexity, coral traits, and reef fish assemblages. *Coral Reefs* 36 (2), 561–575.
- Dennison WC, Barnes DJ, 1988. Effect of water motion on coral photosynthesis and calcification. *J Exp Mar Biol Ecol* 115 (9), 67–77.
- Do ANT, Tran HD, 2023. Combining a deep learning model with an optimization algorithm to detect the dispersal of the early stages of spotted butterflyfish in northern Vietnam under global warming. *Eco. Inform.* 78, 102380.
- Donovan MK, Friedlander AM, Lecky J, Jouffray JB, Williams GJ, Wedding LM, Crowder LB, Erickson AL, Graham NA, Gove JM, Kappel CV, 2018. Combining fish and benthic communities into multiple regimes reveals complex reef dynamics. *Sci. Rep.* 8 (1), 16943. [PubMed: 30446687]
- Dubinsky Z, Stambler N, 1996. Marine pollution and coral reefs. *Glob. Chang. Biol.* 2 (6), 511–526.
- Edmunds PJ, Tsounis G, Lasker HR, 2016. Differential distribution of octocorals and scleractinians around St. John and St. Thomas, US Virgin Islands. *Hydrobiologia* 767, 347–360.
- Fisher WS, 2007. *Stony Coral Rapid Bioassessment Protocol*. US Environmental Protection Agency Office of Research and Development, EPA/600/R-06/167, Washington, DC, 60 pp.
- Fisher WS, 2023. Relating fish populations to coral colony size and complexity. *Ecol. Indic.* 148, 110117, 10 pp.
- Fisher WS, Davis WP, Quarles RL, Patrick J, Campbell JG, Harris PS, Hemmer BL, Parsons M, 2007. Characterizing coral condition using estimates of three-dimensional colony surface area. *Environmental Monitoring and Assessment* 125, 347–360. [PubMed: 17225074]
- Fisher WS, Fore LS, Oliver LM, LoBue C, Quarles RL, Campbell JG, Harris PS, Hemmer BL, Vickery S, Parsons M, Hutchins A, Bernier K, Rodriguez D, Bradley P, 2014. Regional status assessment of stony corals in the U.S. Virgin Islands. *Environ. Monit. Assess.* 186 (11), 7165–7181. 10.1007/s10661-014-3918-z. [PubMed: 25052328]

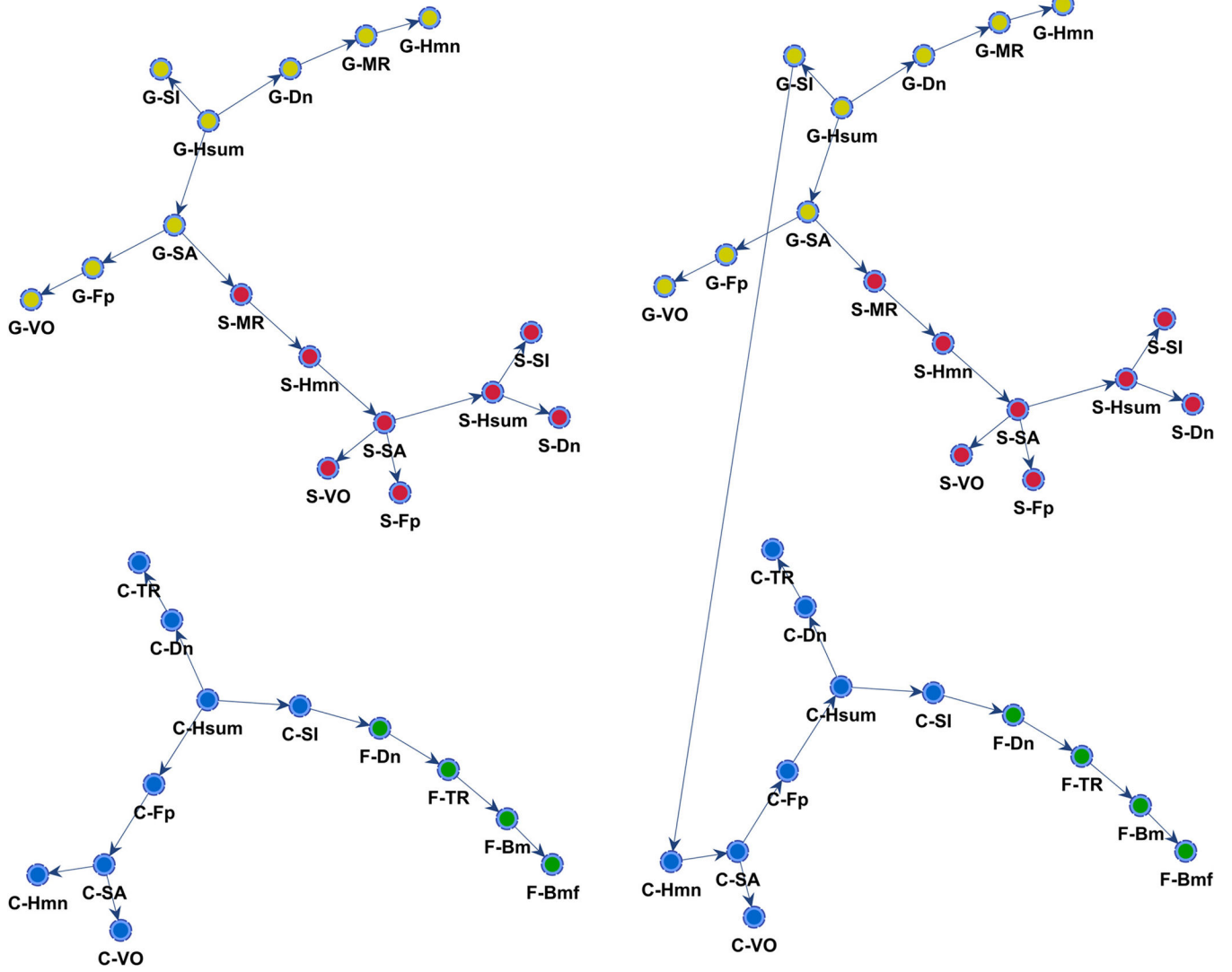
- Fisher WS, Vivian DN, Campbell J, Lobue C, Hemmer RL, Wilkinson S, Harris P, Santavy DL, Parsons M, Bradley P, Humphrey A, Oliver LM, Harwell L, 2019. Regional multi-assemblage status assessment of coral reefs of southern Puerto Rico. *Coast. Manag.* 47, 429–452.
- Ford HV, Gove JM, Davies AJ, Graham NA, Healey JR, Conklin EJ, Williams GJ, 2021. Spatial scaling properties of coral reef benthic communities. *Ecography* 44 (2), 188–198.
- Ford HV, Gove JM, Healey JR, Davies AJ, Graham NA, Williams GJ, 2024. Recurring bleaching events disrupt the spatial properties of coral reef benthic communities across scales. *Remote Sens. Ecol. Conserv.* 10 (1), 39–55.
- Fore LS, Karr JR, Fisher WS, Davis WS, 2008. Making waves with the clean water act. *Science* 322, 1788.
- Fore LS, Karr JR, Fisher WS, Bradley P, Davis WS, 2009. Heeding a call to action for US coral reefs: the untapped potential of the clean water act. *Mar. Pollut. Bull.* 58 (10), 1421–1423. [PubMed: 19615697]
- Froese R, Pauly D, 2007. FishBase. <http://www.fishbase.org>.
- FRRP, 2022. Florida Reef Resilience Network. <https://reefresilience.org/a-collaboration-to-protect-our-coral-reefs/florida/> (accessed Aug 2022).
- Ganesan A, Santhanam SM, 2022. A novel feature descriptor based coral image classification using extreme learning machine with ameliorated chimp optimization algorithm. *Eco. Inform.* 68, 101527.
- Gerassis S, Albuquerque MTD, García JF, Boente C, Giráldez E, Taboada J, Martín JE, 2019. Understanding complex blasting operations: a structural equation model combining Bayesian networks and latent class clustering. *Reliab. Eng. Syst. Saf.* 188, 195–204.
- Gonzalez-Rivero M, Beijbom O, Rodriguez-Ramirez A, Bryant DE, Ganase A, Gonzalez-Marrero Y, Herrera-Reveles A, Kennedy EV, Kim CJ, Lopez-Marcano S, Markey K, 2020. Monitoring of coral reefs using artificial intelligence: a feasible and cost-effective approach. *Remote Sens.* 12 (3), 489.
- Harvell CD, Fenical W, 1989. Chemical and structural defenses of Caribbean gorgonians (*Pseudopterogorgia* spp.): intracolony localization of defense. *Limnol. Oceanogr.* 34, 382–389.
- Hoegh-Guldberg O, Poloczanska ES, Skirving W, Dove S, 2017. Coral reef ecosystems under climate change and ocean acidification. *Front. Mar. Sci.* 4 (158), 1–20.
- Johnson SK, Hallock P, 2020. A review of symbiotic gorgonian research in the western Atlantic and Caribbean with recommendations for future work. *Coral Reefs* 39, 239–258.
- Jones BL, Nordlund LM, Unsworth RKF, Jiddawi NS, Eklof JS, 2021. Seagrass structural traits drive assemblages in small-scale fisheries. *Front. Mar. Sci.* 8, 1–17. 10.3389/fmars.2021.640528. [PubMed: 35685121]
- Kekolahti P, Karikoski J, Riikonen A, 2015. The effect of an individual's age on the perceived importance and usage intensity of communications services—a Bayesian network analysis. *Inf. Syst. Front.* 17, 1313–1333.
- Kelutur JK, Saptarini N, Mustarichie R, Kurnia D, 2021. Bioactive compounds profile of gorgonian corals and their pharmacological activities: a review. *J. Chemother.* 14 (3), 1783–1789. 10.31788/RJC.2021.1436406.
- Kleypas JA, Yates KK, 2009. Coral reefs and ocean acidification. *Oceanography* 22 (4), 108–117.
- Knowlton N, Jackson JBC, 2008. Shifting baselines, local impacts, and global change on coral reefs. *PLoS Biol.* 6, e54 10.1371/journal.pbio.0060054. [PubMed: 18303956]
- Kramer PA, 2003. Synthesis of coral reef health indicators for the western Atlantic: results of the AGRRA program (1997–2000). In: Lang JC (ed) status of coral reefs in the western Atlantic: results of initial surveys, Atlantic and Gulf Rapid Reef Assessment (AGRRA) program Atoll Res. Bull. 496, 1–55.
- Ladd MC, Burkepille DE, Shantz AA, 2019. Near-term impacts of coral restoration on target species, coral reef community structure, and ecological processes. *Restoration Ecol* 27 (5), 1166–1176.
- LaRue EA, Fahey RT, Alveshire BC, Atkins JW, Bhatt P, Buma B, Chen A, Cousins S, Elliott JM, Elmore AJ, Hakkenberg CR, Hardiman BS, Johnson JS, Kashian DM, Koirala A, Papes M, St. Hilaire JB, Surasinghe TD, Zambrano J, Zhai L, Songlin F, 2023a. A theoretical framework for the ecological role of three-dimensional structural diversity. *Front Ecol Environ* 21: 4–13.

- LaRue EA, Knott AJ, Domke GM, Chen HYH, Guo Q, Hisano M, Oswald C, Oswald S, Kong N, Fei S, 2023b. Structural diversity as a reliable and novel predictor for ecosystem productivity. *Front Ecol Environ* 21, 33–39.
- Lenz EA, Bramanti L, Lasker HR, Edmunds PJ, 2015. Long-term variation of octocoral populations in St. John, US Virgin Islands. *Coral Reefs* 34, 1099–1109.
- Lirman D, 1999. Reef fish communities associated with *Acropora palmata*: relationships to benthic attributes. *Bull. Mar. Sci.* 65 (1), 235–252.
- Lumini A, Nanni L, Maguolo G, 2020. Deep learning for plankton and coral classification. *Appl. Comput. Inform.* 19 (3/4), 265–283.
- Maragos JE, Crosby MP, McManus JW, 1996. Coral reefs and biodiversity: a critical and threatened relationship. *Oceanography* 9 (1), 83–89.
- Marre G, Braga CDA, Ienco D, Luque S, Holon F, Deter J, 2020. Deep convolutional neural networks to monitor coralligenous reefs: operationalizing biodiversity and ecological assessment. *Eco. Inform.* 59, 101110.
- Meziere Z, Popovic I, Prata K, Ryan I, Pandolfi J, Riginos C, 2024. Exploring coral speciation: multiple sympatric *Stylophora pistillata* taxa along a divergence continuum on the Great Barrier Reef. *Evol. Appl.* 17 (1), e13644. [PubMed: 38283599]
- Mills MS, Ungeremann M, Rigot G, den Haan J, Leon JX, Schils T, 2023. Assessment of the utility of underwater hyperspectral imaging for surveying and monitoring coral reef ecosystems. *Sci. Rep.* 13 (1), 21103. [PubMed: 38036628]
- Moberg F, Folke C, 1999. Ecological goods and services of coral reef ecosystems. *Ecol. Econ.* 29 (2), 215–233.
- Mora C, 2008. A clear human footprint in the coral reefs of the Caribbean. *Proc. R. Soc. B* 275, 767–773.
- Mumby PJ, Broad K, Brumbaugh DR, Dahlgren CP, Harborne AR, Hastings A, Holmes KE, Kappel CV, Micheli F, Sanchirico JN, 2008. Coral reef habitats as surrogates of species, ecological functions and ecosystem services. *Conserv. Biol.* 22 (4), 941–951. [PubMed: 18477024]
- NOAA, 2014. NOAA Coral Reef Conservation Program National Coral Reef Monitoring Plan. National Oceanic and Atmospheric Administration Coral Reef Conservation Program, Silver Spring, MD. [www.coralreef.noaa.gov](http://www.coralreef.noaa.gov) (accessed Feb 2022).
- Oliver LM, Lehrter JC, Fisher WS, 2011. Relating landscape development intensity to coral reef condition in the watersheds of St. Croix, U.S. Virgin Islands. *Mar. Ecol. Prog. Ser.* 427, 293–302.
- Ozanich E, Thode A, Gerstoft P, Freeman LA, Freeman S, 2021. Deep embedded clustering of coral reef bioacoustics. *J. Acoust. Soc. Am.* 149 (4), 2587–2601. [PubMed: 33940892]
- Pawlik JR, Burch MT, Fenical W, 1987. Patterns of chemical defense among Caribbean gorgonian corals: a preliminary survey. *J. Exp. Mar. Biol. Ecol.* 108, 55–66.
- Pearl J, 1988. *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. Morgan Kaufmann Publishers, Inc, San Francisco, CA.
- Pendleton L. (Ed.), 2008. *The Economic and Market Value of Coasts and Estuaries: What's at Stake?*. Coastal Ocean Values Press, Washington, DC, pp. 135–pp.
- Peterson KD, Evans LC, 2019. Decision support system for mitigating athletic injuries. *Int. J. Comput. Sci. Sport* 18 (1), 45–63.
- Principe PP, Fisher WS, 2018. Spatial distribution of collections yielding marine natural products. *J. Nat. Prod.* 81, 2307–2320. [PubMed: 30299096]
- Principe P, Bradley P, Yee S, Fisher W, Johnson E, Allen P, Campbell D, 2012. *Quantifying Coral Reef Ecosystem Services*. U.S. Environmental Protection Agency Office of Research and Development, EPA/600/R-11/206, Research Triangle Park, NC, 147 pp.
- Reaka-Kudla ML, 2005. Biodiversity of Caribbean coral reefs. In: Miloslavich P., Klein E (Eds.), *Caribbean Marine Biodiversity: The Known and the Unknown*. DEStech Publications, Lancaster, PA, pp. 259–276.
- Roberts CM, McClean CJ, Veron JEN, Hawkins JP, Allen GR, McAllister DE, Mittermeier CG, Schueler FW, Spalding M, Wells F, Vynne C, Werner TB, 2002. Marine biodiversity hotspots and conservation priorities for tropical reefs. *Science* 295, 1280–1284. [PubMed: 11847338]

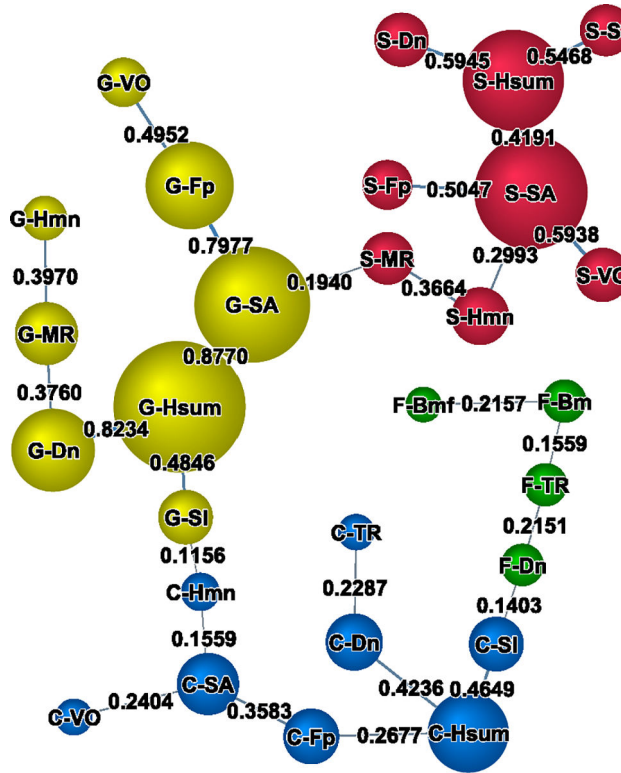
- Rubbens P, Brodie S, Cordier T, Destro Barcellos D, Devos P, Fernandes-Salvador JA, Fincham JJ, Gomes A, Handegard NO, Howell K, Jamet C, 2023. Machine learning in marine ecology: an overview of techniques and applications. *ICES J. Mar. Sci.* 80 (7), 1829–1853.
- Ruzicka RR, Colella M, Porter JW, Morrison JM, 2013. Temporal changes in benthic assemblages on Florida keys reefs 11 years after the 1997/1998 El Niño. *Mar. Ecol. Prog. Ser.* 489, 125–141.
- Santavy DL, Fisher WS, Campbell JG, Quarles RL, 2012. Field Manual for Coral Reef Assessments. US Environmental Protection Agency Office of Research and Development EPA/600/R-12/029 April 2012, Gulf Breeze FL.
- Santavy DL, Jackson SK, Jessup B, Gerritsen J, Rogers C, Fisher WS, Weil E, Szmant A, Cuevas-Miranda D, Walker B, Jeffrey C, Bradley P, Ballantine D, Roberson L, Ruiz-Torres H, Todd B, Smith T, Clark R, Diaz E, Bauza-Ortega J, Horstmann C, Raimondo S, 2022a. A biological condition gradient for coral reefs in the US Caribbean territories: part I. Coral narrative rules. *Ecol. Indic.* 138, 108805.
- Santavy DL, Jackson SK, Jessup B, Horstmann C, Rogers C, Weil E, Szmant A, Cuevas-Miranda D, Walker BK, Jeffrey C, Ballantine D, Fisher WS, Clark R, Ruiz Torres H, Todd B, Raimondo S, 2022b. A biological condition gradient for Caribbean coral reefs: part II. Numeric rules using sessile benthic organisms. *Ecol. Indic.* 135 (7), 108576.
- Schürholz D, Chennu A, 2023. Digitizing the coral reef: machine learning of underwater spectral images enables dense taxonomic mapping of benthic habitats. *Methods Ecol. Evol.* 14 (2), 596–613.
- Sherif FF, Zayed N, Fakhr M, 2015. Discovering Alzheimer genetic biomarkers using Bayesian networks. *Adv. Bioinforma.* 10.1155/2015/639367. Article ID 639367.
- Sorokin YI, 1995. *Coral Reef Ecology*. Ecological Studies, vol. 102. Springer-Verlag e-ISBN-13: 978-3-642-80046-7.
- Spurgeon JPG, 1992. The economic valuation of coral reefs. *Mar Poll Bull* 24 (11), 529–536.
- Sun H, Yue J, Li H, 2022. An image enhancement approach for coral reef fish detection in underwater videos. *Eco. Inform.* 72, 101862.
- Syms C, Jones GP, 2000. Disturbance, habitat structure, and the dynamics of a coral reef fish community. *Ecology* 81, 2714–2729.
- U.S. EPA, 2024. Environmental Dataset Gateway. <https://edg.epa.gov/metadata/catalog/man/home.page>.
- Villon S, Iovan C, Mangeas M, Claverie T, Mouillot D, Villéger S, Vigliola L, 2021. Automatic underwater fish species classification with limited data using few-shot learning. *Eco. Inform.* 63, 101320.
- Woodhead AJ, Hicks CC, Norstrom AV, Williams GJ, Graham NAJ, 2019. Coral reef ecosystem services in the Anthropocene. *Funct. Ecol.* 33, 1023–1034. 10.1111/1365-2435.13331.
- Yee SH, Carriger J, Bradley P, Fisher WS, Dyson B, 2014a. Developing scientific information to support decisions for sustainable ecosystem services. *Ecol. Econ.* 115, 39–50.
- Yee SH, Dittmar JA, Oliver LM, 2014b. Comparison of methods for quantifying reef ecosystem services: a case study mapping services for St. Croix, USVI. *Ecosyst. Serv.* 8, 1–15.



**Fig. 1.** Distance mapping of the coral reef monitoring variables using (a) Pearson correlation and (b) mutual information. Node C-SA is behind C-Fp in 1a and 1b. Coral reef nodes are blue, fish nodes are green, sponge nodes are red and gorgonian nodes are yellow. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

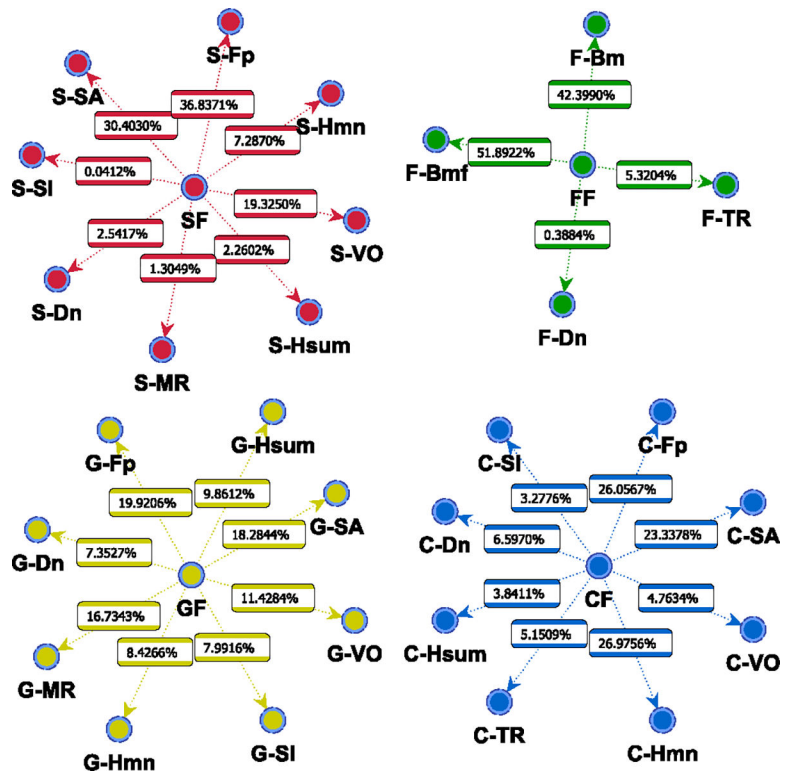


**Fig. 2.** Maximum weight spanning tree structure with (a) structural coefficient of 1.0 and (b) structural coefficient of 0.80 to fully connect all nodes of the network. Coral reef nodes are blue, fish nodes are green, sponge nodes are red and gorgonian nodes are yellow. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

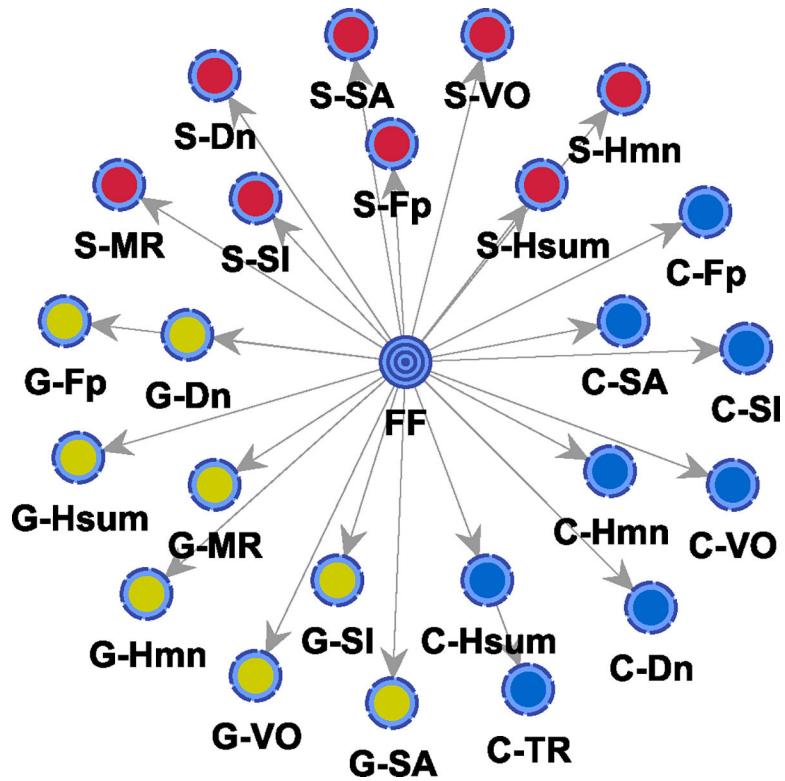


**Fig. 3.** Final Bayesian network structure with node force (node size) and Kullback-Leibler divergence (arc size and numerical values) showing the strength of relationships among measured variables. Coral reef nodes are blue, fish nodes are green, sponge nodes are red and gorgonian nodes are yellow. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

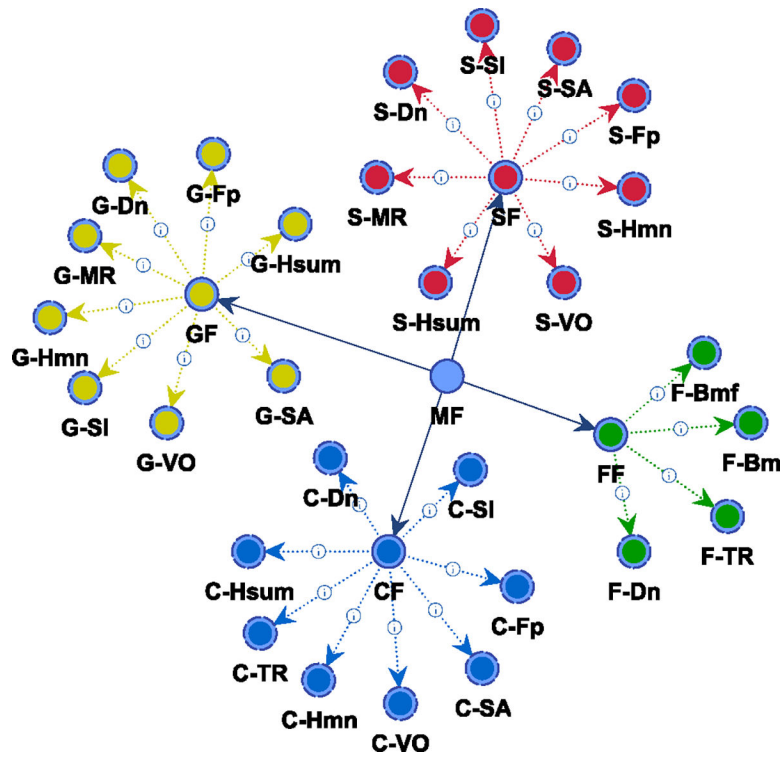




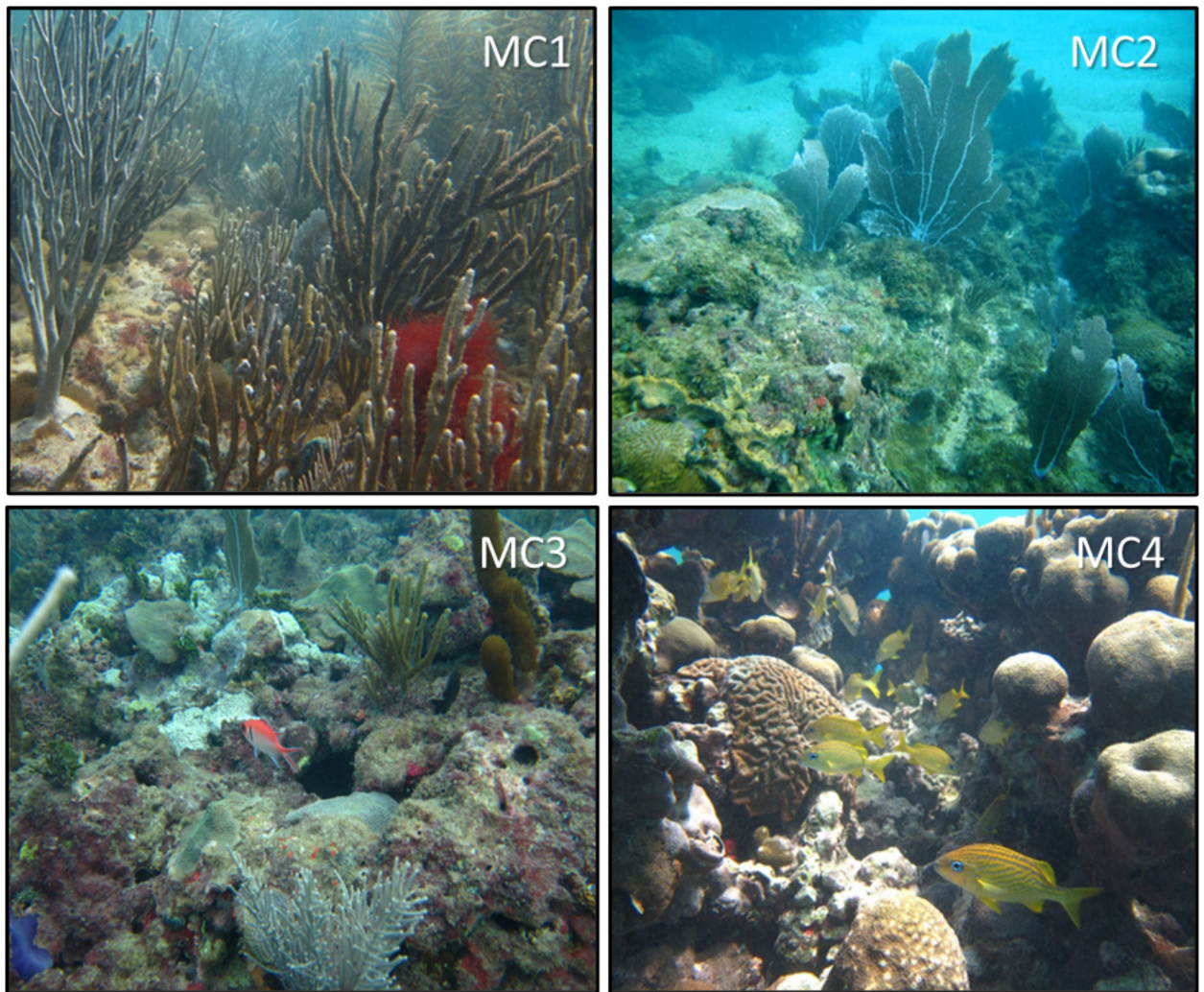
**Fig. 4.** Cluster networks for each community component. Target nodes are latent factors containing cluster states. Percentage numbers are percentage strengths of contributions between the target node and manifest variables. SF = sponge factor; FF = fish factor; GF = gorgonian factor; CF = coral factor.



**Fig. 5.** Network structure for supervised learning of the fish factor (FF) with all measured variables as predictors. The latent FF comprised the manifest variables from corals, sponges and gorgonians.



**Fig. 6.** Hierarchical cluster model structure. The meta-factor (MF) root node clusters the factor nodes for each community component. The leaf (terminal) nodes are the manifest nodes representing probability distributions from measured reef components. Intermediate nodes (between the root and leaf nodes) are the factor nodes for each of the community components. The names of these factor nodes reflect the manifest node with the highest contribution and the number of nodes in the cluster states in parentheses.



**Fig. 7.** Photographs of representative stations for meta-clusters: MC1 (top left), MC2 (top right), MC3 (bottom left) and MC4 (bottom right).

**Table 1**

Posterior mean estimates for sponge variables given each of the identified clusters. Colors are a spectrum from red to green with red, green, and yellow representing lower, higher, and intermediate values, respectively. Distribution of stations (% ,  $n = 138$ ) among sponge clusters are shown in parentheses.

Posterior Mean Values (rows)					
Node	SC1 (9.4%)	SC2 (19.6%)	SC3 (4.3%)	SC4 (24.6%)	SC5 (42.0%)
S-Hmn	11.0297	17.319	18.620	2.377	11.271
S-SA	5938.369	5067.278	12419.130	1234.474	1234.474
S-Hsum	37.592	108.213	176.214	19.953	19.953
S-Fp	2280.440	1038.634	2993.994	237.672	517.318
S-Dn	3.984	5.837	8.300	1.398	1.977
S-SI	53.880	288.020	502.835	53.880	57.870
S-VO	36541.671	19472.920	57656.957	4882.792	4882.792
S-MR	0.752	0.744	0.840	0.201	0.646

**Table 2**

Posterior mean estimates for fish variables given each of the identified clusters. Colors are a spectrum from red to green with red, green, and yellow representing lower, higher, and intermediate values, respectively. Distribution of stations (% , n = 138) among fish clusters are shown in parentheses.

Posterior Mean Values (rows)					
Node	FC1 (34.8%)	FC2 (2.2%)	FC3 (2.9%)	FC4 (50.0%)	FC5 (10.1%)
F-Dn	1.735	1.857	3.388	0.639	1.742
F-Bm	43.716	757.828	169.891	41.031	169.891
F-Bmf	37.089	433.585	34.610	53.575	94.091
F-TR	0.165	0.178	0.217	0.130	0.196

**Table 3**

Posterior mean estimates for gorgonian variables given each of the identified clusters. Colors are a spectrum from red to green with red, green, and yellow representing lower, higher, and intermediate values, respectively. Distribution of stations (% , n = 138) among gorgonian clusters are shown in parentheses.

Posterior Mean Values (rows)					
Node	GC1 (27.5%)	GC2 (14.5%)	GC3 (30.4%)	GC4 (19.6%)	GC5 (7.9%)
G-Hsum	86.742	86.742	312.735	565.793	292.190
G-SA	10330.769	10330.769	39753.851	61838.615	49433.119
G-Dn	3.747	2.630	9.017	14.388	4.836
G-Fp	1665.556	1665.556	5558.179	8566.256	10110.642
G-Hmn	33.633	7.836	34.245	37.620	44.992
G-SI	492.010	226.845	1024.387	1420.144	348.981
G-MR	1.062	0.330	1.413	1.405	1.179
G-VO	142476.082	142476.082	253532.128	611171.884	826475.839

**Table 4**

Posterior mean estimates for coral variables given each of the identified clusters. Colors are a spectrum from red to green with red, green, and yellow representing lower, higher, and intermediate values, respectively. Distribution of stations (% , n = 138) among coral clusters are shown in parentheses.

Posterior Mean Values (rows)					
Node	CC1 (3.6%)	CC2 (8.7%)	CC3 (27.5%)	CC4 (51.4%)	CC5 (8.7%)
C-Hsum	52.963	31.802	36.112	13.131	75.303
C-Dn	2.431	1.690	4.093	1.369	4.425
C-SI	343.133	132.854	162.379	66.733	247.082
C-Fp	10641.328	2871.595	1401.650	1058.027	5410.590
C-SA	83957.155	15835.134	3427.783	3427.783	23279.544
C-TR	0.287	0.367	0.514	0.296	0.532
C-Hmn	33.125	16.955	9.577	10.824	15.111
C-VO	667583.164	187747.084	30155.783	30155.783	256352.700



**Table 5**

Posterior mean estimates for non-fish variables given each of the clusters for the fish factor. Colors are a spectrum from red to green with red, green, and yellow representing lower, higher, and intermediate values, respectively.

Posterior Mean Values (rows)					
Node	FC1	FC2	FC3	FC4	FC5
G-Fp	4153.778	3611.867	1665.556	5927.523	3333.823
G-MR	1.044	0.943	0.969	1.256	0.981
G-Hmn	27.854	37.114	32.553	34.331	31.166
G-VO	229932.718	142476.082	142476.082	421402.360	242426.523
G-SA	26810.019	21131.896	18431.615	39344.322	24535.143
G-Hsum	231.371	162.073	143.240	310.362	219.747
G-Dn	6.604	4.652	5.663	8.433	5.711
G-SI	770.751	674.679	562.720	847.287	621.887
S-Hsum	39.344	45.431	19.953	48.853	59.139
S-MR	0.503	0.502	0.405	0.657	0.473
S-Hmn	8.930	7.631	8.288	11.889	10.873
S-Dn	3.015	2.643	2.332	3.114	3.184
S-Fp	825.149	1156.446	1152.023	843.008	563.278
S-SI	103.633	53.880	53.880	143.222	103.464
S-SA	2700.510	2802.439	2410.448	3164.202	2578.444
S-VO	9572.476	14262.160	11917.318	16252.164	8902.521
C-SI	143.620	156.376	211.798	95.984	167.883
C-Hmn	12.533	9.577	17.878	11.020	15.758
C-TR	0.368	0.336	0.305	0.414	0.305
C-VO	103856.416	30155.783	148349.259	71266.557	97694.912
C-SA	14297.730	3427.783	10872.194	6017.143	7681.732
C-Fp	2600.358	1058.027	2146.168	1688.833	1990.719
C-Dn	2.578	3.719	3.079	2.281	2.408
C-Hsum	29.412	27.938	40.680	26.320	27.161

EPA Author Manuscript

EPA Author Manuscript

EPA Author Manuscript

**Table 6**

Posterior mean estimates for fish variables given each of the clusters for the coral factor (CF). Colors are a spectrum from red to green with red, green, and yellow representing lower, higher, and intermediate values, respectively.

Posterior Mean Values (rows)					
Node	CC1	CC2	CC3	CC4	CC5
F-Dn	2.281	1.627	1.237	1.098	1.245
F-TR	0.181	0.167	0.156	0.144	0.167
F-Bm	92.575	62.508	111.185	59.181	51.770
F-Bmf	146.507	49.481	74.392	56.985	44.524

**Table 7**

Posterior mean estimates for gorgonian variables given each of the clusters for the coral factor (CF). Colors are a spectrum from red to green with red, green, and yellow representing lower, higher, and intermediate values, respectively.

Posterior Mean Values (rows)					
Node	CC1	CC2	CC3	CC4	CC5
G-SI	226.845	407.832	927.441	797.015	881.764
G-MR	0.443	1.195	1.180	1.140	1.241
G-Hmn	17.723	33.693	33.482	30.570	37.334
G-Dn	2.630	4.652	7.809	8.188	5.691
G-Hsum	86.742	143.240	282.750	287.385	279.580
G-Fp	1665.556	2638.712	5185.779	5194.390	5558.179
G-SA	10330.769	18431.615	32968.972	35207.242	38073.744
G-VO	142476.082	220215.314	352750.627	320712.514	414563.395

**Table 8**

Posterior mean estimates for all manifest variables given each of the meta-clusters (MCs) for the meta-factor node. Colors are a spectrum from red to green with red, green, and yellow representing lower, higher, and intermediate values, respectively. Distribution of stations (% , n = 138) among meta-clusters are shown in parentheses.

Posterior Mean Values (rows)				
Node	MC1 (44.9%)	MC2 (35.5%)	MC3 (8.0%)	MC4 (11.6%)
G-Hsum	419.955	155.924	86.742	126.548
G-SA	50776.465	19337.835	10330.769	17057.505
G-Fp	7528.990	2857.175	1665.556	2964.481
G-Dn	10.749	4.996	3.747	3.934
G-MR	1.376	0.930	1.062	0.915
G-VO	492447.722	176472.831	142476.082	234917.054
G-SI	1098.690	568.398	492.010	441.113
G-Hmn	37.275	25.397	33.633	28.642
S-Hmn	13.438	8.696	9.169	6.075
S-Hsum	63.059	22.473	60.071	39.486
S-SI	167.664	56.160	160.308	110.748
S-Dn	3.921	2.098	3.416	2.369
S-SA	3771.308	1906.459	2976.658	2632.556
S-MR	0.682	0.534	0.448	0.364
S-Fp	1019.180	689.294	601.745	634.646
S-VO	16528.482	9405.489	11514.669	11479.563
C-Fp	1648.591	1149.192	2269.386	6093.073
C-SA	5749.159	3427.783	9744.253	39449.644
C-Hsum	27.164	19.228	36.019	52.009
C-VO	57129.397	30155.783	108025.067	359135.114
C-Hmn	11.034	10.493	12.876	21.432
C-SI	117.173	92.109	150.648	234.262
C-Dn	2.554	2.092	3.001	2.776
C-TR	0.391	0.354	0.442	0.393
F-Dn	0.904	1.517	0.850	1.943
F-TR	0.141	0.162	0.140	0.177
F-Bm	51.856	90.176	117.909	83.146
F-Bmf	54.184	63.362	91.805	47.467