# Five practical uses of spatial autocorrelation for studies of coral reef ecology

S Hamylton
*University of Wollongong*, shamylto@uow.edu.au

# Five practical uses of spatial autocorrelation for studies of coral reef ecology

## Abstract

The organisation of benthic communities across coral reefs is underpinned by spatially structured ecological processes and neighbourhood interactions such as larval dispersal, migration, competition and the spread of disease. These give rise to spatial autocorrelation in reef communities. This paper demonstrates how the measurement of spatial autocorrelation can profitably be incorporated into studies of coral reef ecology through a series of 5 simple statistical exercises: for the generation of maps depicting the strength of spatial relationships between ecological communities, as an indicator of optimal dimensions for sampling ecological communities on coral reefs, as a diagnostic tool for model misspecification, as an indicator of a spatial process underpinning the distribution of an observed community pattern and as a surrogate for missing variables in a model. The benefits of incorporating spatial autocorrelation include (1) quantifying the extent and pattern of autocorrelation across reefs, (2) signifying the presence of redundant information in field datasets, (3) indexing the nature and degree to which fundamental assumptions of classic (i.e. non-spatial) statistical techniques are violated, (4) indicating the nature (spatial versus non-spatial) of an observable pattern to be modelled, and (5) offering an opportunity to partition out and utilise the spatially structured component of model error as a surrogate for a missing variable. Collectively, the statistical exercises presented here provide a persuasive case for the measurement and interrogation of spatial autocorrelation in studies of coral reef ecology.

## Keywords

## Disciplines

Medicine and Health Sciences | Social and Behavioral Sciences

## Publication Details

**Why use spatial statistics in coral reef ecology? Five practical reasons to measure spatial autocorrelation**

S. Hamylton and V.Harwood

*School of Earth and Environmental Sciences, University of Wollongong, Australia, NSW2522.*

Email: sarah_hamylton@uow.edu.au

Tel: +61 42213589

## Abstract

The organisation of benthic communities across coral reefs is underpinned by spatially structured ecological processes and neighbourhood interactions such as larval dispersal, migration, competition and the spread of disease. These give rise to spatial autocorrelation in reef communities. This paper demonstrates how the measurement of spatial autocorrelation can profitably be incorporated into studies of coral reef ecology through a series of five simple statistical exercises: for the generation of maps depicting the strength of spatial relationships between ecological communities, as an indicator of optimal dimensions for sampling ecological communities on coral reefs, as a diagnostic tool for model misspecification, as an indicator of a spatial process underpinning the distribution of an observed community pattern and as a surrogate for missing variables in a model. The benefits of incorporating spatial autocorrelation include i. quantifying the extent and pattern of autocorrelation across reefs, ii. signifying the presence of redundant information in field datasets, iii. indexing the nature and degree to which fundamental assumptions of classic (i.e. non-spatial) statistical techniques are violated, iv. indicating the nature (spatial vs. non-spatial) of an observable pattern to be modelled, and v. offering an opportunity to partition out and utilise spatially structured components of model error as a surrogate for a missing

variable. Collectively, the statistical exercises presented here provide a persuasive case for the measurement and interrogation of spatial autocorrelation in studies of coral reef ecology.

*Spatial statistics, Autocorrelation, Moran, Geary*

## 1.0 Introduction

It has long been established that coral reefs and the benthic communities they support, including hard and soft corals, sponge populations, algae and seagrasses, are structured in space (Vaughan, 1915; Done, 1983). This structure arises from contemporary environmental influences such as water depth, exposure to incident waves, ambient light availability, hurricane activity and terrestrial input from rivers (Geister, 1977; Sheppard, 1982; Kleypas 1999; Harmelin-Vivien, 1994; Fabricius, 2005a), biological influences such as competition, larval dispersal and the spread of disease (Paris-Limouzy, 2012; Porter et al., 2001) and geological influences, such as the configuration of antecedent Pleistocene platform and availability of suitable substrate for the settlement of coral larvae (Hopley et al. 2007; Hubbard, 1997). These factors, which often act in synergy across the reef seas, lead to the presence of spatial autocorrelation in coral reef communities. Spatial autocorrelation refers to the correlation of a single reef characteristic as a function of its position in geographic space, such that characteristics at proximate locations tend to be related. This is a fundamental property of most ecological datasets collected on coral reefs. It arises because of ecological processes that abide by Tobler's First Law of Geography: *"everything is related to everything else, but near things are more related than distant things"* (Tobler, 1970).

Positive spatial autocorrelation means that geographically nearby reef characteristics, such as percentage live coral cover, tend to be similar because of spatially structured processes and neighbourhood interactions (Hamylton, 2012a). The presence of spatial autocorrelation introduces numerous deviations from the assumptions of classical statistics that warrant attention in studies of

coral reef ecology. These deviations can be addressed through the application of spatial statistics, a collection of analytical techniques and models in which a clear association is maintained and exploited between quantitative data and the spatial coordinates that locate them (Chorley 1972). The breadth of disciplinary interest in spatial analysis is evidenced by several review papers, books and edited collections on the subject, spanning ecology marine metapopulations, rainforest ecology, urban ecology and reserve design (Legendre, 1993; Lichstein et al., 2002; Fortin and Dale, 2005 McIntire and Fajardo, 2009)). The development of geospatial technology in the form of geographical information systems, geographical positioning systems and remote sensing instruments over the last fifty years has provided exciting opportunities for the analysis of spatial patterns on coral reefs. Yet many studies of marine ecology, in particular those of coral reefs, fail to adequately address spatial autocorrelation (but for examples of spatially explicit studies of coral reef ecology, see Table 1 of Hamylton et al., 2012b; of the 11 studies listed, only 1 explicitly incorporates spatial autocorrelation).

This paper demonstrates the practical value of incorporating spatial autocorrelation into coral reef ecological studies through a series of five simple statistical exercises:

(1) to depict the strength of spatial relationships between ecological communities across a given geographical area,

(2) as an indicator of optimal dimensions for sampling ecological communities on coral reefs,

(3) as a tool for diagnosing model misspecification,

(4) as an indicator of a spatial process underpinning the distribution of an observed community pattern, and

(5) as a surrogate for missing variables in a model.

In relation to the first exercise, coral reef ecologists often wish to detect and characterise spatial patterns across a reef platform to ascertain how community assemblages are organised along gradients (Bak and Newland, 1995; Fabricius et al. 2005b). To achieve this, spatial autocorrelation is measured at both the global and local scales to provide information on the nature of spatial relationships between benthic communities across a complete reef platform.

In relation to the second exercise, the collection of community composition information in the field relies on an appropriate sampling strategy given the inherent spatial variability of the community (Fortin and Dale, 2005). This may vary across reef zones and attempts have been made to define optimal sampling approaches for coral reef fishes (Houk et al., 2006), hard corals (English et al., 1997; Murdoch and Aronson, 1999) and soft sediment benthic communities (Schlacher et al., 1998). In accordance with Tobler's First Law of Geography, if two points on a coral reef are close together in space, it follows that they will be similar in character. Depending on distance between samples, the presence of spatial autocorrelation can therefore indicate information redundancy where a field campaign has sampled points that are close together. The second exercise derives a semivariogram from multiple measures of autocorrelation to specify the distance between two field locations at which the variance between two points is no longer distance-dependent and they can be considered independent of each other.

The third, fourth and fifth exercises draw links between community patterns and environmental processes, which is a common objective for marine ecologists (Vellend, 2010). As large-scale georeferenced datasets, sophisticated statistical methods and adequate computing power have become increasingly available, coral reef studies more frequently employ model-based statistical inference to determine whether spatial variation in community composition can be explained by

4

environmental factors (Harborne et al., 2006; Mellin et al., 2010; Arias-Gonzalez et al., 2011; Pittman and Brown, 2011). Spatial autocorrelation can be usefully employed in such studies (for example, see Mellin et al. 2010). This utility is demonstrated in the third exercise, where autocorrelation is used to diagnose model misspecification for a common modelling scenario that utilises a classic statistical approach (e.g. ordinary least squares regression) as opposed to spatial statistical approach.

The presence of interactive (or neighbourhood-context) effects in ecological communities suggests a need for a model with a spatially dependent covariance structure (Cliff and Ord 1981). The fourth exercise achieves this by introducing an spatially explicit autoregressive term to the ordinary least dquares equation. This has the effect of regressing the dependent variable against values of itself at a given distance away (or spatial lag). The fifth exercise subdivides the error term associated with the regression equation into *spatially structured unexplained* and *unexplained* components. This enables the spatial structure of error to be built into the model without the cause necessarily being known. The approaches outlined in these latter exercises avoid statistical pitfalls associated with failing to account for autocorrelation in ecological datasets and draw on the inherent spatial structure of the data to enhance model performance.

## 1.1 Study Site and Datasets employed

The study site selected for these exercises was Central Bommie, a lagoonal platform reef at One Tree island on the Great Barrier Reef (Figure 1). This reef platform (area 7066 $m^2$) sits in approximately 2.5 m water depth, the shallow surface of central platform becomes explosed at low tide and is topped by a matrix of unconsolidated carbonate, algae and sparse corals. The deeper platform periphery and sides remain submerged throughout the tidal cycle and are composed of a

concentric ring of live coral, soft corals and dead coral on which calcified algae have encrusted. This configuration is typical of many intertidal lagoonal reefs, both on the Great Barrier Reef and elsewhere in the reef seas due to the differing degrees of aerial exposure of platform surfaces (Hopley et al. 2007). Figure 1 illustrates the associated datasets utilized for the statistical exercises in the present study, which include a satellite image, a digital elevation model and a detailed record of the benthic community in the form of an underwater phototransect across the reef platform.
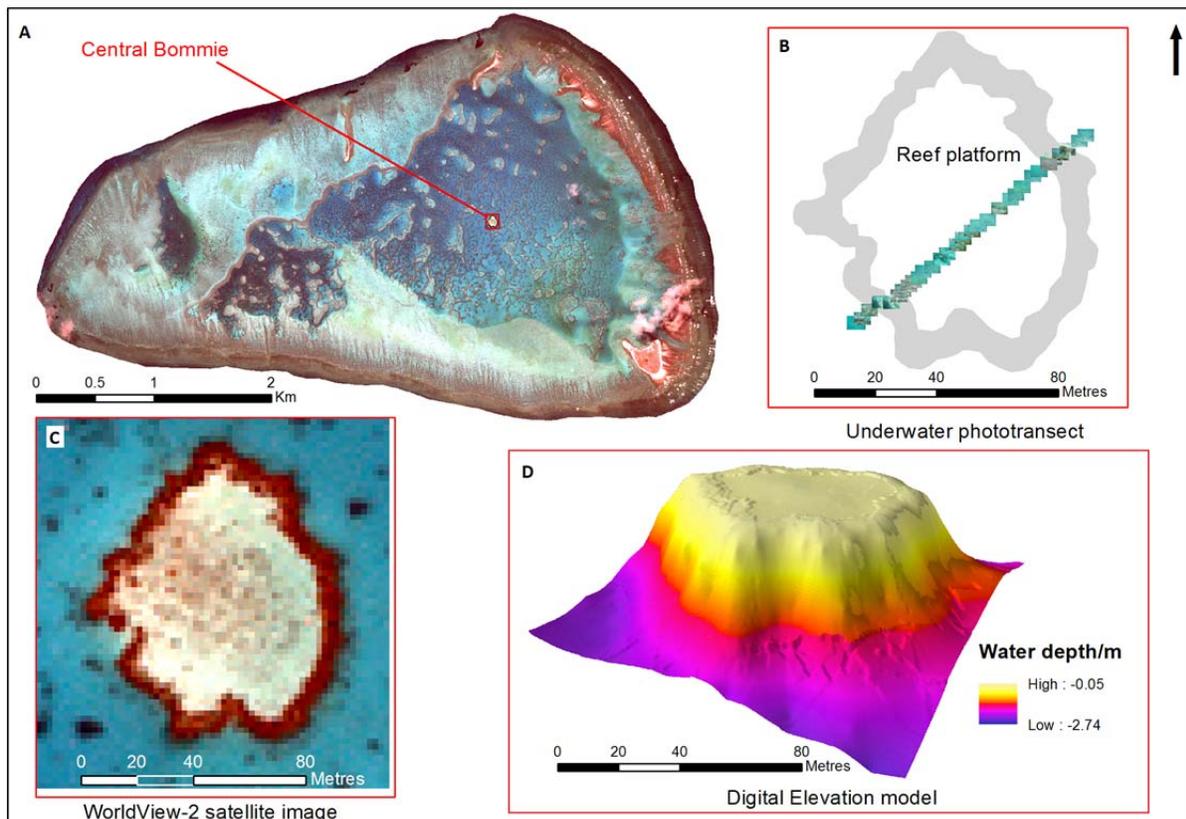


**Figure 1**. A Location of Central Bommie (152°4'43.20"E; 23°29'48.68"S), the case study site at One Tree Island and B-D the datasets associated with the present study. B A transect profile of 47 underwater photographs across the reef platform, C WorldView-2 satellite image of Central Bommie, D A digital elevation model of Central Bommie.

Table 1 and Figure 1 summarise the datasets collected to support the statistical exercise on Central Bommie. These include a WorldView-2 satellite image covering the entire island and reef system, a collection of detailed underwater photographs along a transect profile across the reef platform from

which data on community assemblages were derived, and two datasets of bathymetric point

measurements of water depth from which a detailed bathymetric model of the reef platform structure

was derived.

| Dataset | Description |
| --- | --- |
| WorldView-2 satellite image of One Tree Island and reef system acquired in October 2011 | Satellite image composed of 8 wavebands, spatial resolution 2 m. Image was processed to retrieve benthic community reflectance through the application of atmospheric and water column correction (see section 2.1). |
| Detailed underwater photographs taken in October 2011 | 42 underwater photographs taken *in-situ* across a transect traversing the reef platform with a Nikon D7000 digital SLR camera in a Nauticam housing. Photographs were taken approximately 1 m apart, 50 cm above the ground such that each image covered a field of view of approximately 1 $m^2$ with an overlap of approximately 20% on either side of the image. |
| Community composition data | Each underwater photograph was visually assessed and the presence and % cover of the following community components were recorded: live branching coral, live encrusting coral, dead coral, carbonate sand, rubble, macroalgae, calcified algae, sponges and invertebrates. These multivariate measures of community composition were then reduced to a single variable using the ordination technique of correspondence analysis. |
| Laser Airborne Depth Sounder (LADS) bathymetric point measurements | 52 water depth points acquired using an airborne Laser Airborne Depth Sounder (LADS) |
| Ceeducer Pro echosounder bathymetric point measurements | 105 water depth points measured with a Ceeducer Pro single beam echosounder mounted to the hull of a boat. All depth measurements across the reef platform were taken during October 2011 and corrected to mean sea level using data from a tide gauge installed nearby at Shark Alley in the One Tree Island lagoon. The tide gauge was composed of a pressure transducer and Campbell 21X datalogger. |
| Digital bathymetric model | A synoptic digital bathymetric model of the reef platform was derived by interpolating the |

| | 157 water depth points to a continuous raster surface of 1m spatial resolution. This was achieved by applying an inverse distance weighting algorithm to the shapefile dataset of 157 bathymetric point measurements. |
|---|---|

**Table 1.** The datasets employed by the case studies presented.

## 2.0 Methodology

### 2.1 A map depicting the strength of spatial relationships between ecological communities across a study area

To quantify spatial autocorrelation it was necessary to regress the correspondence for multiple measures of reef community character against the distance between point locations for which that character was measured in space. The satellite image was pre-processed to correct for the influence of the atmosphere and water column on light transfer using standard image processing algorithms (Cooley et al. 2002; Lee et al. 1998, 1999). Reflectance values could then be interpreted as indicative of benthic community character. The Moran's $I$ statistic was used to capture the extent to which the reflectance in the blue band of the Worldview2 image (450-500 nm wavelength) covaried with itself across space. This was calculated as the cross product for a given reflectance value, z ,at location $i$ across a defined neighbourhood, N:

$$I\,(d) = \frac{\frac{1}{W}\sum_{(i)}\sum_{j\in N(i)} w_{ij}(z_{(i)} - \tilde{z})(z_{(j)} - \tilde{z})}{\frac{1}{n}\sum_{j\in N(i)}(z_{(i)} - \tilde{z})^2}$$

Equation 1

where   $d$ = distance class on which Moran's $I$ is calculated

   $z_i$ = benthic community reflectance at location i

   N = the neighbourhood within which reflectance values are sampled

   $\tilde{z}$ =  the mean of the z values

   $W$= the sum of the weights $w_{ij}$ for the given distance class.

Moran's *I* takes the value of 1 when sites *i* and *j* are at or within a distance *d* and 0 otherwise. In this way, only the pairs of sites (*i,j*) within the stated distance class (*d*) of each point location were taken into account. This yielded a large and positive statistic in the presence of positive spatial dependence, a large and negative statistic in the presence of negative spatial dependence and was close to zero with a random map pattern.

The statistic was calculated at both the local and global scale for benthic reflectance values represented in a raster grid using the Spatial Autocorrelation (Moran's I) tool in ArcGIS 10. For the local scale calculation, a neighbourhood of 20 m was defined around each pixel in the raster grid and comparisons were drawn against all pixels falling inside this. At the global scale, each pixel location was systematically compared to every other pixel location in the dataset.

**2.2 An indicator of optimal dimensions for sampling ecological communities on coral reefs**

A semivariogram was generated using the exploratory function within the Geostatisical Analyst tools of ArcGIS10. This was generated using values from the preprocessed raster grid file that corresponded to benthic community reflectance (see section 2.1). The semivariogram was based on the Geary's *c* statistic, which measured correspondence for a given distance class *d*, on the basis of the squared difference of a particular characteristic between two point locations $z_i$ and $z_j$:

$$c(d) = \frac{\frac{1}{2W}\sum_{(i)}\sum_{j \in N(i)} w_{ij}(z_{(i)} - z_{(j)})^2}{\frac{1}{n-1}\sum_{j \in N(i)}(z_{(i)} - z_{(j)})^2}$$

Equation 2

This measure ranged from 0 to some unspecified value larger than 1. It was small and positive if the map pattern has positive spatial dependence, large and positive if the map pattern had negative spatial dependence and intermediate between these extremes if the map pattern was random, with a value of 1 under the null hypothesis of no spatial correlation. The corresponding semivariogram

went beyond immediate neighbours to decompose and describe spatial structure for a given series of spatial lag (distance) classes (through first, second and third order neighbours) as follows:

$$\gamma(h) = \left(\frac{1}{N}(h)\right) \sum_{i} \sum_{(j|d(i,j)=h)} (z_{(i)} - z_{(j)})^2 \qquad\qquad \text{Equation 3}$$

where $h$ = the order of neighbours defined (e.g. third order).

## 2.3 A tool for diagnosing model misspecification

An ordinary least squares model was run to regress community composition as a response variable against water depth as an independent variable using the freely available spatial analysis software GeoDa (Anselin, 2003). Community composition was estimated as a dependent variable for each of 42 underwater photographs taken along transect across the reef platform and represented in a spatial dataset as a point shapefile (Table 1). Photographs were taken approximately 1 m apart, 50 cm above the ground such that each image covered a field of view of approximately 1 m$^2$ with an overlap of approximately 20% on either side of the image. Each photograph was visually interpreted and the percentage area covered was estimated for the following categories: live branching coral, live encrusting coral, dead coral, carbonate sand, rubble, macroalgae, calcified algae, sponges and invertebrates. This generated a multivariate dataset that captured a wide range of reef community types. The ordination technique of correspondence analysis (Legendre and Legendre, 1998) was then used to reduce the multivariate dataset in a manner that best represented the multi-dimensional nature of ecological data (Dray et al. 2012). Values from the first correspondence axis, which explained 88% of the variability in the dataset, were taken as a measure of community composition and treated as the response variable.

The average water depth of the area of reef platform covered by each underwater photograph was extracted from the digital bathymetric model (Table 1). This information was added to the point shapefile using the Add Surface Information tool within the 3d Analyst tools of ArcGIS10. The

ordinary least squares regression was therefore run using GeoDa on a shapefile composed of 42 points along with their corresponding community and water depth information across the reef platform. Residuals of the ordinary least squares model were calculated for each point and to diagnose the presence or absence of spatial autocorrelation, the Moran's *I* statistic of the model residuals was calculated.

## 2.4 An indicator of a spatial process underpinning the distribution of an observed community pattern

The ordinary least squares regression model described in section 2.3 was respecified as an autoregressive model to incorporate a neighbourhood context effect operating through a spatially lagged expression of the response variable itself:

$$\mu_{(i)} = \beta_0 + \beta_1 X_{1(i)} + \rho \sum_{j \in N(i)} w_{ij} Y(j) + e_{(i)} \qquad i = 1, ..., n \qquad \text{Equation 4}$$

where     $\rho$ = a parameter associated with the interaction effect,

n = the number of sample locations,

$X_{1(i)}$ is the independent variable (in this case, water depth) at location *i*,

$Y(j)$ is the community composition at location *j*

e(i) = independent, normally distributed error term,

$\beta_0$ and $\beta_1$ = coefficients estimated using the model.

To introduce a spatially lagged autoregressive term, it was necessary to construct a spatial weights matrix ($w_{ij}$), to express for each case those locations that belonged to its neighbourhood, such that $w_{ij}$=1 when i and j were neighbours and $w_{ij}$=0 otherwise (Anselin and Bera 1998). A range of different weights matrices can be constructed to incorporate varying definitions of the neighbourhood surrounding a point (e.g. within a user-defined Euclidean distance band, or

11

selecting a specified number of nearest neighbours). The spatial weights matrix and the spatially

lagged autoregressive model were constructed in GeoDa using the <span style="color:red">shapefile of 47 points with</span>

<span style="color:red">corresponding coral cover and water depth information</span> described in section 2.3.

## 2.5 A surrogate for a missing model variable

Finally, a spatial error model was constructed using the same community dataset as that described

in sections 2.3 and 2.4. It assumed that the unexplained variation was normally distributed and

partitioned out the spatially structured component of model error, which was expressed alongside a

spatial autoregressive component as follows:

$$Y(i) = \beta_0 + \beta_1 X_1(i) + u(i) \hspace{4cm} \text{Equation 5}$$

$$u(i) = \rho \Sigma_{j \in N(i)} w(i,j) u(j) + e(i) \hspace{3cm} \text{Equation 6}$$

where     $u(i)$ = sum error of the linear regression model for case i,.

$u(j)$ = sum error of the linear regression model for case j,

$e(j)$ = unexplained random error.

As with the autoregressive model described in section 2.4, <span style="color:red">the spatial weights matrix and the</span>

<span style="color:red">spatial error model were constructed in GeoDa using the shapefile of 47 points with corresponding</span>

<span style="color:red">community composition and water depth information.</span>

**3.0 Results**

**3.1 A map depicting the strength of spatial relationships between ecological communities across a study area**

Figure 2 illustrates the Moran's *I* statistic of spatial autocorrelation, calculated locally around each point for a neighbourhood radius of 20 m (plotted across the raster grid) and also calculated globally (inset univariate Moran scatterplot). At the local scale, low values indicated negative spatial autocorrelation was apparent in a concentric ring around the reef platform periphery, with higher values across the upper platform surface. At the global (i.e. reef platform) scale, positive spatial autocorrelation was apparent (Moran's I = 0.90).
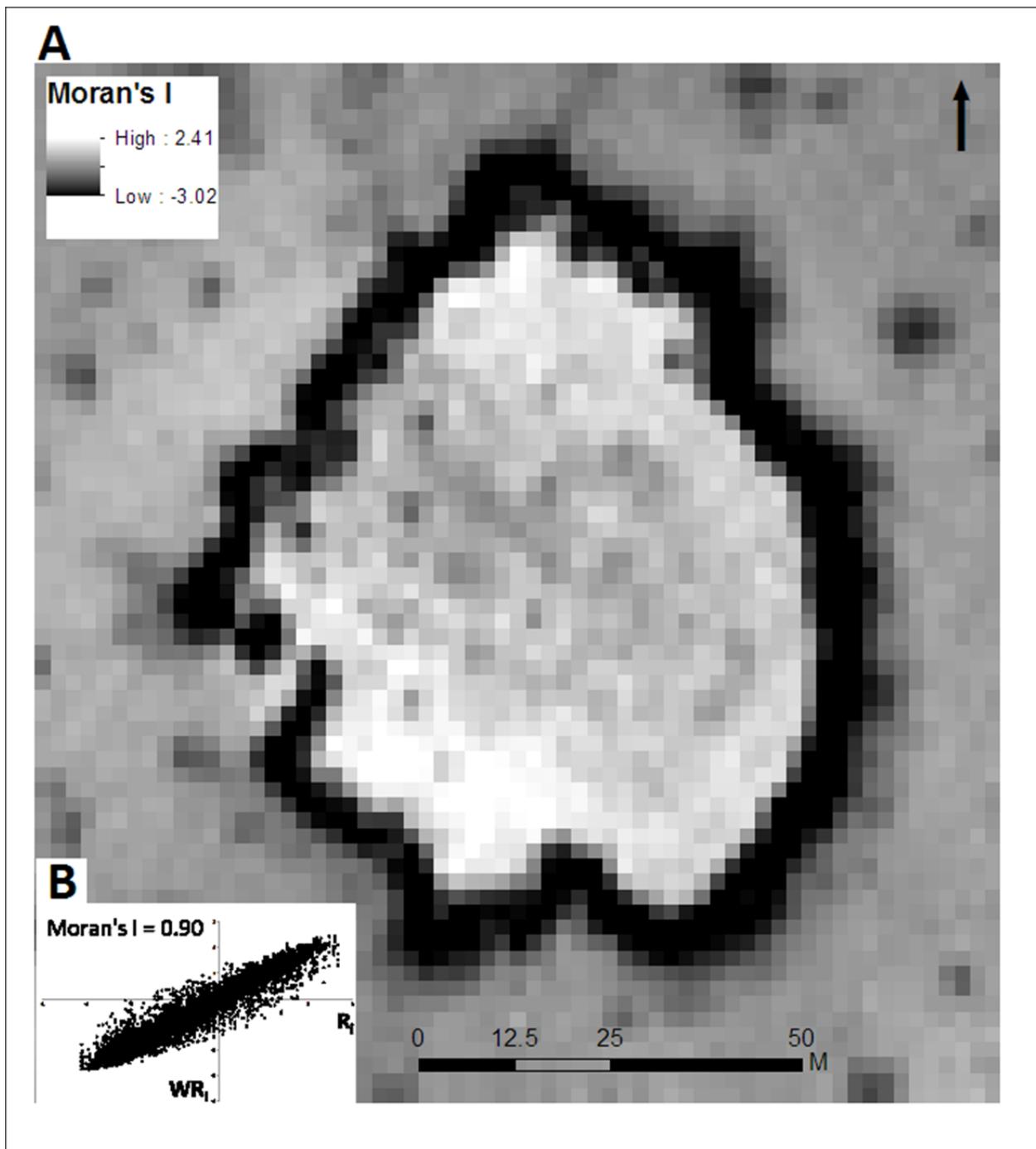
**Figure 2.** Spatial autocorrelation (Moran's I) of the reflectance values that comprise the satellite image of the reef platform at One Tree Island (spatial resolution 1 m). A. Measured locally for each point in the raster grid (within a neighbourhood of 20 m), and B. Measured globally by comparing each point systematically to every other point in the dataset. The univariate Moran scatterplot shows the spatial lag of the variable (reflectance) on the y-axis ($WR_1$) and the original variable ($R_1$) on the x-axis .

## 3.2 An indicator of optimal dimensions for sampling ecological communities on coral reefs

Figure 3 illustrates a plot of the spatial lag between two point locations against the Geary statistic, which yielded a curve with a sill point, a nugget and a range. The sill point represented the point at which Geary's $c$ levelled off and no further increase in the statistic was observed as the distance between point pairs increased. This corresponded to a given distance range beyond which benthic community components no longer influenced each other. The nugget represented the value of the Geary statistic at distance $= 0$ m. In the case of the reflectance values of the reef platform satellite image, the semivariogram was best described (that is, with minimal residuals) by a Gaussian function with a sill at $\gamma = 1.217 \times 10^3$ and a range of approximately 20 m.
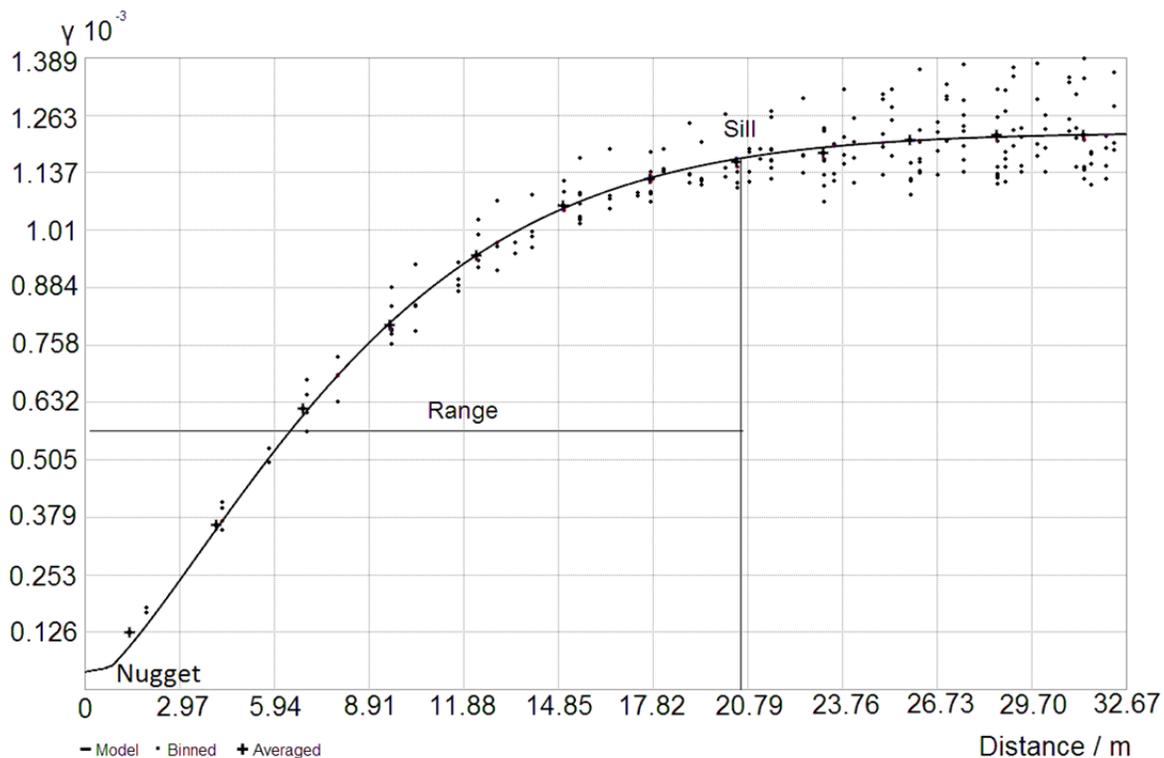


**Figure 3.** Semivariogram for the reflectance values of the reef platform satellite image, best described with a Gaussian function with a range of approximately 20 m (distance between which the characteristics associated with point locations on the reef platform no longer influence each other), a nugget of $0.06 \times 10^{-3}$ and a sill at $\gamma = 1.217 \times 10^{-3}$.

**3.3 A tool for diagnosing model misspecification, an indicator of a spatial process**

**underpinning the distribution of an observed community pattern and a surrogate for a**

**missing model variable**

Table 2 summarises the results of three regression models of community composition across

Central Bommie against water depth across the reef platform. In the first instance, an ordinary least

squares regression was run, in the second instance, a spatially lagged autoregressive model was run

and in the third instance, a spatial error model was run. In the second and third models,

incorporation of the autoregressive terms enhanced model performance ($R^2$ increased from 0.87 for

the ordinary least squares model to 0.93 and 0.91 respectively for the spatially lagged

autoregressive model and the spatial error model). In addition to this, the presence of spatial

dependence in the model residuals was reduced (Moran's *I* reduced from 0.93 to -0.07 and -0.01).

| ORDINARY LEAST SQUARES REGRESSION | |
|---|---|
| **Adjusted $R^2$** | 0.87 |
| **N (degrees freedom)** | 42 (40) |
| **Moran's I (residuals)** | 0.93 |
| **β Coefficient** | -8.64 |
| **β Water depth** | -93.50 |
| **T-statistic Coefficient (p value)** | -3.44 (p < 0.002) |
| **T-statistic Water depth (p value)** | -16.33 (p<0.001) |
| **SPATIALLY LAGGED AUTOREGRESSION** | |
| **Adjusted $R^2$** | 0.93 |
| **Moran's I (residuals)** | -0.07 |
| **β Coefficient** | -4.28 |
| **β Water depth** | -42.28 |
| ***P* interaction parameter** | 0.57 |
| **SPATIAL ERROR REGRESSION** | |
| **Adjusted $R^2$** | 0.91 |
| **Moran's I (residuals)** | -0.01 |
| **β Coefficient** | -7.21 |
| **β Water depth** | -86.83 |
| ***P* interaction parameter** | 0.44 |

**Table 2.** Results of a classic ordinary least squares regression (section 2.3), spatially lagged
autoregression (section 2.4) and spatial error regression (section 2.5) between the reef community
composition on the reef platform and water depth. $R^2$ indicates proportion of variation observed in

16

the dataset explained by the regression equations, β corresponds to the coefficient applied in the regression equation and the p-values associated with the T-test indicate the probability of attaining the result observed under the null hypothesis.

## 4.0 Discussion

The magnitude and spatial configuration of locally- measured autocorrelation plotted across the map in Figure 2 indicated the variable extent to which Tobler's Law is being played out across the reef surface (Haining, 2003). As has been observed in terrestrial ecosystems, the measures of Moran's I indicated contradictory processes, i.e. both positive and negative autocorrelation, at different locations and scales (Van de Koppel et al. 2006). The observed pattern indicated that proximity between two points on the reef surface was likely to induce similarity between the positively autocorrelated areas of carbonate sand deposition on the platform surface, and dissimilarity between the negatively autocorrelated hard and soft coral and algal communities around the platform rim. This negative autocorrelation could result from self organisation of community components into a pattern that could reflect, for example, competition for light (Lang, 1990). In contrast to the raw reflectance values depicted in the satellite image, this statistic is informative on the nature (attractive vs repulsive) of relationships between the benthic community at each location and that within its neighbourhood across the reef platform.

It should be noted that the value of the Moran's $I$ statistic depends on the size of the spatial units employed for its calculation. While too fine sampling units may result in noisy spatial correlation patterns, too large units may exaggeratedly smooth out spatial structures. Where possible, the size of sampling units should be selected with *a priori* regard for the spatial scales at which potential underlying influences on benthic community structure are manifest (Fortin and Dale, 2005). The semivariogram (section 2.2, Figure 3) gives an indication of suitable dimensions for this purpose. In cases where it is not possible to compute this, cross scale analysis provides a useful guide to the

17

appropriate spatial scales for studying ecological phenomena, both in terrestrial systems (Dray et al. 2012; Kiel et al. 2012) and on coral reefs (Gust et al. 2001). This approach is particularly useful given the scale-dependent spatial variability of coral reef assemblages, such as hard coral communities (Murdoch and Aronson, 1999).

In accordance with the sill point in the semivariogram (Figure 3), a sampling scheme devised to optimise information capture and reduce redundancy across this coral reef platform would space sample points at least 20 m apart. This was identified as the distance between two field locations at which the variance between two points was no longer distance-dependent and they could be considered independent of each other. Such an observation has important implications when considering whether to collect field data from a boat-based platform, SCUBA or snorkel survey. Furthermore, the relative remoteness of many reefs and the logistical challenges associated with sampling the benthic communities inhabiting them, in particular the need to utilise SCUBA equipment, often leads to the geographical concentration of field data collection effort. This observation is also therefore useful for identifying information redundancy in sample points that are close together and consequently similar. Observed patterns of spatial variability have been found to be scale-dependent on coral reefs because of the interaction of multiple forces operating at different scales (Edmunds and Bruno, 1996; Hughes et al. 1999). Extrapolation across scales that were not sampled can therefore be problematic because observations made for a given sample scale may not hold at other scales. In relation to this issue, the semivariogram approach outlined here can be thought of as spanning spatial scales that range from the greatest to the smallest distance between points sampled (i.e. from 1 to 33 m). The transferability of this approach is also contingent on consistency in large scale influences on coral reef ecology such as hurricane influence (Edmunds and Bruno, 1996).

It would appear at first sight that there is a strong inverse relationship between the community composition on the reef and water depth ($R^2 = 0.87$), which may be underpinned by the inability of many reef community components such as coral to tolerate aerial exposure during low phases of the tidal cycle (Anthony and Kerswell, 2007). However, the Moran's *I* value of 0.93 computed on the residuals is diagnostic of spatial dependence, which violates assumptions of classic ordinary least squares regression. The first assumption is that observations are independent of each other and the second is that residuals are both normally distributed and randomly located. In the presence of spatial dependence, the effective number of degrees of freedom in the sample is smaller than the one estimated from the number of observations. This is because proximate observations are not independent of each other and cannot be freely permuted at random to create the reference (null) distribution of the test statistic. As a consequence, statistical tests of model significance generate narrow confidence limits. Regressing autocorrelated data cases in an ordinary least squares model may therefore increase the likelihood of a Type I error, inflate the goodness of fit measure and underestimate the standard error as a result of allocating some of the effect due to interaction to the existing dependent variables (Cliff and Ord 1981). Many reef studies that employ regression analysis do not report any assessment of residuals, leaving these assumptions untested. This warrants scrutiny because of the spatial structure observable on reefs due to contemporary environmental, biological and geological influences (Done, 2012).

While the presence of autocorrelation in model residuals is indicative of a statistical pitfall, spatially referenced residuals can be mapped to provide a useful clue as to the distribution and underlying nature of missing covariates. Where positive residuals cluster together on the map, the tendency for the model to overestimate community characteristics may indicate the need for an additional covariate that has the overall effect of reducing the predicted characteristic in this geographic area,

and vice versa. Hamylton and Spencer (2011) provide an example of this spatially explicit technique for exploring model performance.

Improvements in model performance and reductions in spatial dependence of residuals with the incorporation of the autoregressive term (section 2.4) suggested that the assumptions of observation independence and random error were being held in a more robust manner (Lichstein et al. 2002). Where there is reason to specify neighbourhood interaction, spatial regression is preferable to classic regression and easily achievable using freely available software such as GeoDa (or alternatively, R or Spatial Analysis in Macroecology).

The practical value of fitting a spatial error model (section 2.5) is that the spatially structured component of the error can effectively be partitioned out and eliminated from the residuals, thereby patching the model so that valid inferences can be drawn from the predictors (Haining, 2003). This draws explicitly on the information held in the residuals about the behaviour of the response variable. By adopting this approach, the analyst maintains faith with the original set of predictors and keeps these in the model, whilst partitioning out the residuals into stochastic and spatially correlated components. Although the spatially correlated components can be modelled and explained in a statistical sense, their identity remains unknown. It is possible, however, to infer that they arise from neighbourhood-context interactions (Cliff and Ord, 1981). This approach to model development originates from a simple, well defined initial model and progresses toward a more general model by adding autocorrelation parameters. As is evident from the comparison with the ordinary least squares model, this enhances the power of the regression (on this occasion, $R^2$ increased from 0.87 for the ordinary least squares model to 0.91). This is valuable for applications that seek to derive continuous predictions of coral reef community characteristics across large

areas (e.g. Hamylton et al. 2012b) to assist with the spatial planning of marine reserves (Roberts et al., 2003; Sobel and Dahlgren, 2004; Almany et al., 2009).

The approaches outlined in 2.4 (invoking a spatially lagged autoregressive component) and 2.5 (modelling the spatially structured component of the unexplained model variation) both add a spatially explicit term to the regression equation. The decision of whether to select an autoregressive model or a spatial error model lies largely with the modelling scenario that faces the analyst. Autoregression provides a statistically robust approach where there is reason to believe that the response variable might influence itself through a neighbourhood context effect. For example, the characteristic of hard coral cover might be positively spatially autocorrelated because the location of a spawning coral might influence the settlement site of its offspring because of the interconnected dynamics of larval dispersal (Paris-Limouzy, 2012). In contrast, a spatial error model may be more appropriate in a modelling scenario where the analyst wishes to retain the original set of predictors without adding additional independent variables. This situation might occur when all known theoretical influences, or at least those that can in practice be conceptualised and represented digitally, have been accounted for and incorporated into the model specification.

## 5.0 Concluding remarks

Studies of coral reef ecology frequently employ statistical inference, the dependability of which is based upon the validity of assumptions about how ecological processes play out across reefs. While the statistical techniques demonstrated in this paper have been widely applied by ecologists in other environments (Dray et al. 2012), evidence of their application to coral reefs is limited (but see Mellin et al. 2010, Hamylton 2012a, Hamylton et al. 2012b). The collective statistical exercises presented here provide a persuasive case for the measurement and interrogation of spatial

autocorrelation in coral reef studies. Firstly, spatial autocorrelation enables us to quantify the extent and spatial patterning realised by the application of Tobler's First Law of Geography to coral reefs. It provides information on the benthic community at a given location on a reef platform is related to its neighbourhood. This is particularly useful when combined with remotely sensed datasets that provide synoptic information on community reflectance in an accurate and consistent manner. Secondly, spatial autocorrelation identifies and quantifies the extent of redundant information in field datasets, indicating the optimal dimensions of sampling schemes. Third, spatial autocorrelation indexes the nature and degree to which a fundamental statistical assumption is violated, and, in turn, indicates the extent to which conventional statistical inferences are compromised when non-zero spatial autocorrelation is overlooked. Autocorrelation complicates statistical analysis by altering the variance of spatially distributed information, increasing the risk of making incorrect statistical decisions (e.g., positive spatial autocorrelation results in an increased tendency to reject the null hypothesis when it is true). With the proliferation of modelling exercises that seek to explain and predict aspects of marine ecology, this warrants scrutiny if our understanding of the processes we study is to develop correctly. Fourth, the incorporation of a spatially explicit component to regression models is instructive on the nature of an observable spatial pattern to be modelled, suggesting the appropriate approach to be used (spatial vs. non-spatial). Fifth, spatial autocorrelation draws on explainable components of model error as a surrogate for a missing variable, enabling us to enhance model performance. This is achieved by partitioning out and utilising structured components of model error as a surrogate for a missing variable. Without spatial autocorrelation, the ecological character of coral reefs would exhibit a limited geographic expression and appear completely random; with it, it exhibits spatial organization, which is the hallmark of many shallow benthic coral reef communities.

## Acknowledgements

## References

Almany GR, Connolly SR, Heath DD, Hogan JD, Jones GP, McCook LJ, Mills M, Pressey RL, Williamson DH (2009) Connectivity, biodiversity conservation and the design of marine reserve networks for coral reefs. Coral Reefs 28:339–351

Anselin, L (2003) GeoDa 0.9 User's Guide. Spatial Analysis Laboratory (SAL). Department of Agricultural and Consumer Economics, University of Illinois, Urbana-Champaign, IL.

Anselin L, Bera A (1998) Spatial dependence in linear regression models with an introduction to spatial econometrics. In: Ullah A, Giles DE (eds) Handbook of applied economic statistics, Marcel Dekker Press, New York

Anthony KRN and Kerswell AP (2007) Coral mortality following extreme low tides and high solar radiation. Marine Biology 151:1623-1631

Arias-Gonzalez JE, Acosta-Gonzalez G, Membrillo M, Garza-Perez JR, Castro-Perez JM (2011) Predicting spatially explicit coral reef fish abundance, richness and Shannon-Weaver index from habitat characteristics. Biodivers Conserv 21: 115−130

Bak RPM, Nieuwland G (1995) Long-Term Change in Coral Communities Along Depth Gradients Over Leeward Reefs in the Netherlands Antilles. Bulletin of Marine Science 56:609-619

Chorley RJ (1972) Spatial analysis in geomorphology. In: Chorley RJ (ed) Spatial Analysis in Geomorphology. Methuen & Co. Ltd, UK

Cliff AD, Ord JK (1981) Spatial processes: models & applications. Pion, London

Cooley, T (2002) FLAASH, a MODTRAN-4 based atmospheric correction algorithm, its application and validation. Proceedings of the geoscience and remote sensing symposium 3: 1414–1418Dray, S, Pélissier, R, Couteron, P, Fortin, MJ, Legendre, P, Peres-Neto, PR, Bellier,E, Bivand, R,, Blanchet, FG,, De Cáceres, M, Dufour, AB, Heegaard, E,, Jombart T, Munoz, F, Oksanen, J. Thioulouse, J Wagner. HH (2012) Community ecology in the age of multivariate multiscale spatial analysis. Ecological Monographs 82:  257-275

Done TJ (2012) Corals: Environmental Controls on Growth. In: Hopley D (ed) The Encyclopaedia of Modern Coral Reefs. Springer, BerlinEdmunds PJ, Bruno, JF (1996) The importance of sampling scale in ecology: kilometer-wide variation in coral reef communities. Mar Ecol Prog Ser 143: 165-171

English S, Wilkinson C, Baker V (1997) Survey Manual for Tropical Marine Resources, 2nd Edition. Australian Institute of Marine Science, Townsville

Fabricius, K (2005a) Effects of terrestrial runoff on the ecology of corals and coral reefs: review and synthesis. Marine Pollution Bulletin 50: 125-146

Fabricius K, De'ath G, McCook L, Turak E, McWilliams D (2005b) Changes in algal, coral and fish assemblages along water quality gradients on the inshore Great Barrier Reef. Marine Pollution Bulletin 51:384-398

Fortin MJ, Dale M (2005) Spatial Analysis: A guide for ecologists. Cambridge University Press, Cambridge

Geister J (1977) The influence of wave exposure on the ecological zonation of Caribbean coral reefs. In: Taylor DL (ed) Proc 3rd International Coral Reef Symposium Vol. 2: Geology. Rosenstiel School of Marine and Atmospheric Science, Miami, Florida

Gust, N, Choat, J, Howard, J, Mccormick, MI (2001) Spatial variability in reef fish distribution, abundance, size and biomass : a multi-scale analysis. Marine Ecology Progress Series 214:237-251.

Haining RH (2003) Spatial Data Analysis: Theory and Practice. Cambridge University Press, Cambridge

Hamylton S, Spencer T (2011) Geomorphological modelling of tropical marine landscapes: Optical remote sensing, patches and spatial statistics. Continental Shelf Research 31:151-161

Hamylton S (2012a) A comparison of spatially explicit and classic regression modelling of live coral cover using hyperspectral remote sensing data in the Al Wajh lagoon, Red Sea. International Journal of Geographic Information Science 26: 2161-2175

Hamylton S, Spencer T, Hagan A (2012b) Spatial modeling of benthic cover using remote sensing data in the Aldabra lagoon, Western Indian Ocean. Mar Ecol Prog Ser 460: 35-47

Harborne AR, Mumby PJ, Pychaluk K, Hedley JD, Blackwell PG (2006) Modeling the beta diversity of coral reefs. Ecology 87: 2871−2881

Harmelin-Vivien, ML (1994) The effects of storms and cyclones on coral reefs: A review. J. Coastal Res 12: 211–231.

Hopley D, Smithers S, Parnell K (2007) The Geomorphology of the Great Barrier Reef: Development, Diversity and Change. Cambridge University Press, Cambridge

Houk P, Van Woesik R (2006) Coral reef benthic video surveys facilitate long-term monitoring in the Northern Mariana Islands: toward an optimal sampling strategy. Pacific Science 60:117-189

Hubbard DK (1997) Reefs as Dynamic Systems, pp 43 – 67, In: Birkeland, C. (Ed.) (1997) The Life and Death of Coral Reefs. Chapman and Hall, UK

Hughes TP, Baird, AH Dinsdale, EA Moltschaniwskyj, NA Pratchett, MS, Tanner, JE, Willis, BL (1999) Patterns of recruitment and abundance of corals along the Great Barrier Reef. Nature 397:59-62

Keil, P, Schweiger,O, Kühn, I, Kunin, WE, Kuussaari, M, Settele, J, Henle, K, Brotons, L, Pe'er, G, Lengyel, S, Moustakas, A, Steinicke, H, Storch, D (2012) Patterns of beta diversity in Europe: the role of climate, land cover and distance across scales. Journal of Biogeography. 39, 1473–1486

Kleypas J, McManus J, Menez L (1999) Environmental limits to coral reef development: where do we draw the line? American Zoologist 39:146-159

Lang JC, Chornesky EA (1990) Competition between Scleractinian reef corals – a review of mechanisms and effects. In: Dubinsky Z (ed) Ecosystems of the World, 25. Coral Reefs. Elsevier Science Publishing Company, Amsterdam, The Netherlands

Lee, ZP, Carder, KL, Mobley, C, Steward, RG, Patch, JS  (1998) Hyperspectral remote sensing for shallow waters. 1. A semi-analytical model. Applied Optics 37: 6329–6338

Lee, ZP, Carder, KL, Mobley, C, Steward, RG, Patch, JS  (1999) Hyperspectral remote sensing for shallow waters. 2. Deriving bottom depths and water properties by optimization. Applied Optics 38: 3831–3843

Legendre P (1993) Spatial Autocorrelation: Trouble or new paradigm? *Ecology,* 74: 1659-1673.

Legendre P, Legendre LFJ (1998) Numerical Ecology. Elsevier, Amsterdam

Lichstein JW, Simons TR, Shriner SA, Franzreb KE (2002) Spatial Autocorrelation and autoregressive models in ecology. Ecological Monographs 73:445-463

McIntyre, EJB, Fajardo, A (2009) Beyond description: the active and effective way to infer processes from spatial patterns. Ecology 90:46–56

Mellin C, Bradshaw C, Meekan MG, Caley MJ (2010) Environmental and spatial predictors of species richness and abundance in coral reef fishes. Glob Ecol Biogeogr 19: 212−222Murdoch TJT, Aronson, RB (1999)  Scale-dependent spatial variability of coral assemblages along the Florida Reef Tract. Coral Reefs 18: 341-351

Paris-Limouzy CB (2012) Reef Interconnectivity/ Larval Dispersal, pp.881-889. In: Hopley D (ed) The Encyclopaedia of Modern Coral Reefs. Springer, Berlin

Porter JW, Dustan P, Jaap WC, Patterson KL, Kosmynin V, Meier OW, Patterson E, Parsons M (2001) Patterns of spread of coral disease in the Florida Keys. Hydrobiologia 460:1-24

Pittman SJ, Brown KA (2011) Multi-scale approach for predicting fish species distributions across coral reef seascapes. PLoS ONE 6: e20583

Roberts C, Andelman S, Branch GR, Bustamante JC, Castilla J, Dugan B, Halpern H, Leslie K, Lafferty J, Lubchenco D, McArdle, D Possingham H, Ruckleshaus M, Warner R (2003) Ecological criteria for evaluating candidate sites for marine reserves. Ecological Applications 13:199-214

Schlacher TA, Newell P, Clavier J, Schlacher-Hoenlnger J, Chevillon C, Britton J (1998) Soft sediment benthic community structure in a coral reef lagoon - the prominence of spatial heterogeneity and 'spot endemism'. Mar Ecol Prog Ser 174:159-174

Sheppard CRC (1982) Coral populations on reef slopes and their major controls. Mar Ecol Prog Ser 7:83-115

Sobel J, Dahlgren C (2004) Marine reserves: a guide to science, design and use. Island Press, Washington DC

Tobler W (1970) A computer movie simulating urban growth in the Detroit region. Economic Geography 46:234–40

Van de Koppel, J, Altieri, AH, Silliman, BR, Bruno, JF, Bertness, MD (2006) Scale-dependent interactions and community structure on cobble beaches. Ecology Letters 9:45–50


Vaughan TW (1915) On recent Madreporaria of Florida, the Bahamas, and the West Indies, and on collections from Murray Island, Australia.  Yearbook of the Carnegie Institute of Washington 14:220- 231

Vellend, M (2010) Conceptual synthesis in community ecology. The Quarterly review of biology 85:1830206