

Analysing the influence of African dust storms on the prevalence of coral disease in the Caribbean Sea using remote sensing and association rule data mining

Heather Hunter & Guido Cervone

To cite this article: Heather Hunter & Guido Cervone (2017) Analysing the influence of African dust storms on the prevalence of coral disease in the Caribbean Sea using remote sensing and association rule data mining, International Journal of Remote Sensing, 38:6, 1494-1521

To link to this article: <http://dx.doi.org/10.1080/01431161.2016.1277279>

 Published online: 31 Jan 2017.

 Submit your article to this journal [↗](#)

 View related articles [↗](#)

 View Crossmark data [↗](#)



Analysing the influence of African dust storms on the prevalence of coral disease in the Caribbean Sea using remote sensing and association rule data mining

Heather Hunter ^a and Guido Cervone ^{b,c}

^aDepartment of Applied Marine Physics, Rosenstiel School of Marine and Atmospheric Science, Miami, FL, USA; ^bDepartment of Geography and Institute for CyberScience, Pennsylvania State University, University Park, PA, USA; ^cLamont-Doherty Earth Observatory, Columbia University, Palisades, NY, USA

ABSTRACT

The application of an association rule data mining algorithm is described to combine remote sensing and *in-situ* geophysical data to show a relationship between African dust storms, Caribbean climate, and Caribbean coral disease. An analysis is performed to quantify the relative statistical significance of each Caribbean climate parameter on the prevalence of coral disease. Results show that African dust storms contribute to an increased prevalence of coral disease in the Caribbean Sea, and that the correlation between them is influenced by other climate parameters, especially sea surface temperature.

ARTICLE HISTORY

Received 29 July 2016

Accepted 22 December 2016

1. Introduction

About 11% of the historical extent of coral reefs around the world has been lost since the early 2000s, and an additional 16% is severely damaged (Gardner et al. 2003). In the Caribbean Sea, alone, approximately 90–95% of major reef-building corals have suffered mass mortalities in the last forty years (Garrison et al. 2003). Coral reefs in this region cover thousands of kilometres of coastline and play an integral role in the region's economy as a source of food, shoreline protection, and biopharmaceuticals (Burke and Maidens 2004). Degradation to the coral, however, is associated with reduction in abundance and diversity of fish, as well as a decline in topographical complexity (Schutte, Selig, and Bruno 2010). Research shows that the key drivers of this degradation include coastal development, sediment run-off, over-fishing, and pollution (Burke and Maidens 2004), as well as increasing sea surface temperatures (SSTs) and disease outbreaks.

Most coral evolved to thrive in a water temperature range between 23° and 29°C, and deviations from this range are known to induce thermal stress (Raymundo, Couch, and Harvell 2008; Shinn et al. 2000). The photosynthetic processes of the corals' symbiotic algae breakdown when the SST deviates from normal maximum or minimum by as little as 1.0°C over a continuous period of 2–3 days (Burke and Maidens 2004). These anomalies trigger the

CONTACT Heather Hunter  hhunter@miami.edu  Department of Applied Marine Physics, University of Miami, Miami, FL, USA

expulsion of the corals' photosynthetic symbionts, resulting in the appearance of white pigmentation known as coral bleaching (Eakin et al. 2010). In recent years, both SST anomalies and observations of coral bleaching have increased in the Caribbean significantly (Eakin et al. 2010; Baker, Glynn, and Riegl 2008).

Aeolian dust transported from the Saharan region of Africa is believed to be contributing to the increased incidence of coral disease in the Caribbean over the last forty years (Garrison et al. 2003). The entrainment process that lifts the dust into the troposphere also captures bacteria, fungi, and other microorganisms (Griffin and Kellogg 2004). The dust itself contains nutrients vital to marine ecosystems such as nitrogen, phosphorous, and iron; however, coral has evolved to thrive in nutrient-limited environments (Griffin and Kellogg 2004). When there is an abundance of nutrients, particularly iron, sea-surface dwelling phytoplankton spawns harmful blooms that impede incoming solar radiation and weaken coral immunity (Griffin and Kellogg 2004). Observations by Selig et al. (2006) show a shift from coral- to algal-dominated communities in the last forty years, implying a shift in nutrient availability, which has occurred in tandem with increasing incidence of disease. Bruno et al. (2007) and Shinn et al. (2000) also suggest a relationship between SST anomalies and coral disease outbreaks, supported by small-scale field studies, in addition to a relationship between coral disease and other climate variables. The mere presence of pathogens in dust does not guarantee coral succumb to disease, but the resulting physiological stress due to the combined effects of enhanced dust transport, nutrient enrichment, and SST anomalies encourages the proliferation of disease (Shinn et al. 2000).

Previous work (Schutte, Selig, and Bruno 2010) has identified the need for quantifying spatio-temporal changes in Caribbean coral communities at regional scales. Schutte, Selig, and Bruno (2010) conducted *in-situ* surveys to quantify the relationship between coral decline and macroalgal cover. Field sampling has its limitations in which such sampling requires a large amount of data across a large spatial region that can be costly and labour intensive (Nurlidiasari and Budhiman 2005); however, combining field data with satellite data can provide additional insight not previously explored in depth. With the increasing availability of high-resolution satellite observations, it is possible to perform a comprehensive assessment of coral reefs, environmental change, and the response of coral to such changes at larger spatial scales. Since most documented coral disease outbreaks occur at spatial scales of entire ocean basins (Selig et al. 2006), satellite remote sensing is well-suited to monitor the variation of coral disease incidence with changes in the regional environment. For example, high-resolution optical imagery is already used to monitor visual changes in diverse tropical environments (Chauvaud, Bouchon, and Maniere 1998) and Florida coral reef communities (Palandro et al. 2003).

Andréfouët and Riegl (2004) present a thorough assessment of the use of remote sensing in studying coral reefs, from aerial photography to satellite observation. In particular, satellite observations of aerosol optical depth in optical wavelengths by the moderate-resolution imaging spectroradiometer (MODIS) (Prospero et al. 2002) and the cloud-aerosol lidar and infrared pathfinder satellite observation (CALIPSO) (Adams, Prospero, and Zhang 2012) show that Saharan dust events are derived from multiple source points that vary seasonally. CALIPSO data show that the magnitude and extent of Saharan storms also vary seasonally, with annual variations primarily driven by the North Atlantic Oscillation (NAO), coupled with the Arctic Oscillation and the El Niño Southern

Oscillations (ENSO) (Garrison et al. 2003; Griffin and Kellogg 2004; Goudie and Middleton 2001). During positive phases of the NAO (associated with above-average temperatures in the Eastern USA) and periods of strong ENSO, the Sahara experiences drier than normal conditions, and the flux of dust is at a maximum. This correlation between maximum dust transport and the ENSO was directly observed in the average surface concentration of African dust measured *in situ* at Barbados during the 1983–1984 ENSO (Garrison et al. 2003). In recent years, satellite imagery has increasingly been used to track African dust storms across the Atlantic.

This study uses an Association Rule Data Mining (ARDM) algorithm on a combination of satellite remote-sensing and *in-situ* data to show a relationship between Saharan dust storms, Caribbean climate, and the prevalence of coral disease, previously hypothesized in studies that used direct observations or surveys (Selig et al. 2006; Shinn et al. 2000; Bruno et al. 2007; Schutte, Selig, and Bruno 2010). In addition, the relative statistical significance of each climate parameter is estimated to show that the prevalence of coral disease is not only impacted by the presence of Saharan dust, but also by other environmental parameters known to exhibit an influence on coral reef health (Selig et al. 2006). The satellite remote-sensing measurements selected include SST, ultraviolet (UV) absorbing aerosol index (AAI), the diffuse light attenuation coefficient at 490 nanometres (nm) (K490), and chlorophyll-*a* concentration (chl-*a*). The ENSO Southern Oscillation Index (SOI) was also used in this study. The analysis of these parameters over a period of 13 years, paired with *in-situ* observations of coral reef health, provides a deeper understanding of the environmental states that drive change in the Caribbean coral community and also shows the relationship between Saharan dust and coral disease.

2. Background information

2.1. Coral disease

In general, coral disease is ‘any impairment to health resulting in physiological dysfunction’ (Raymundo, Couch, and Harvell 2008). It involves interactions between a host, an agent, and the environment. In coral, disease is typically manifested as a change in morphology or colour, e.g. lesion growth or white discoloration (Raymundo, Couch, and Harvell 2008). To date, only a few of the total number of infectious biotic diseases observed in coral in the Caribbean have identified pathogens, and the precise reasons for the rapid spread of these diseases in the last few decades remain a poorly understood issue. Pathogens have many supposed vectors and reservoirs, and some disease may be associated with change in corals’ microbial communities, not necessarily an infectious pathogen (Selig et al. 2006). Indeed, drivers of coral disease are not necessarily new pathogens but may be other environmental processes that contribute to physiological stress, such as poor water quality and increased SST (Raymundo, Couch, and Harvell 2008). In fact, in the Caribbean, the increase in frequency and severity of coral disease is coincident with major changes in structure and function as a result of anthropogenic and natural causes (Raymundo, Couch, and Harvell 2008). These and other stressors in the climate may act synergistically to promote disease outbreaks.

In the 1970s, an outbreak of white band disease resulted in widespread mortality of staghorn coral (Miller et al. 2009). In 1983, observations showed Caribbean-wide coral reef mortalities and the infection of Caribbean Sea fans (Shinn et al. 2000). In 1987, mass coral bleaching occurred in conjunction with a warm summer, and the proliferation of the coral reef black band disease (Shinn et al. 2000). In the mid-1990s, the proliferation of aspergilliosis, a disease affecting sea-fans caused by the soil fungus *aspergillus*, was reported (Shinn et al. 2000). Because the fungus is not known to reproduce in seawater, and there were no fungal spores found in the Caribbean Sea, these observations led to the hypothesis that there must be some external, periodic source of soil (Garrison et al. 2003).

Observations have also shown that coral bleaching is an ongoing process exacerbated during warm summer months, especially in years of strong ENSO. Unfortunately, many coral pathogens grow optimally between 30°C and 35°C, which is within the anomalous temperature range for Caribbean coral (Miller et al. 2009). In the Caribbean Sea during the 1980s and 1990s, annual coral bleaching was observed to increase logarithmically with SST anomalies; a 0.1°C increase in local SST caused a corresponding 35% increase in the extent of reported coral bleaching, and local SST increases of 0.2°C resulted in mass coral bleaching events (Baker, Glynn, and Riegl 2008). As observations of SST anomalies have continued increasing in recent years, observations of coral disease have also increased, leading to the hypothesis that high SST may promote growth and reproduction of pathogens (Eakin et al. 2010; Baker, Glynn, and Riegl 2008; Miller et al. 2009). In 2005, a large thermal anomaly resulted in a mass-bleaching event in the Caribbean (Eakin et al. 2010), and this mass-bleaching event was correlated with outbreaks of white plague and yellow-band disease (Miller et al. 2009).

2.2. Influence of Saharan dust

Dust from North Africa to the Caribbean Sea is transported westward between 5° and 25°N along easterly trade winds with a southern boundary limited by the intertropical convergence zone, and a transit time of approximately 1 week (Moulin et al. 1997; Adams, Prospero, and Zhang 2012). Inspection of satellite measurements from MODIS (Prospero et al. 2002) and CALIPSO (Adams, Prospero, and Zhang 2012) suggests that these North African dust events are derived from multiple source points. The precise source of the dust varies seasonally. That is, during warmer months – May–September – the primary source of African dust is near the Niger Delta region in Mali, while in colder months – October–April – the primary source is near Lake Chad (Garrison et al. 2003).

In the Caribbean, *in-situ* measurements of microbe distribution in air samples have indicated that both bacteria and fungi cross the Atlantic from North Africa with the desert dust (Griffin and Kellogg 2004). When African dust is not present, the concentration of bacteria and fungi in Caribbean air samples is approximately half that measured when the African dust is present (Garrison et al. 2003). Since the late 1970s, coincident with accelerated coral reef decline, the amount of transatlantic dust available in the atmosphere above the Caribbean Sea has increased (Prospero and Nees 1986).

African dust is primarily comprised quartz, clay minerals, iron oxides, and calcium carbonate (Gasso, Grassian, and Miller 2010). It also contains major nutrients, such as phosphorus and iron (Shinn et al. 2000). Among all of these elements, the most

significant for coral reefs is iron. Once deposited in the ocean, iron feeds surface-dwelling phytoplankton, but an excessive amount of iron results in algal blooms that can inhibit sunlight and enable opportunistic marine microbial pathogens previously held in check by nutrient limitations to infect the coral. If deposited dust particles are sufficiently small, the particles' residence time in the photic zone increases, and over such long time-scales, iron can be extracted from the dust, maintaining high nutrient levels (Garrison et al. 2003; Gasso, Grassian, and Miller 2010).

As discussed in Section 1, it is hypothesized that Saharan dust plays a significant role in the increasing incidence of coral disease in the Caribbean, primarily due to the transport of nutrients and bacteria. It is suspected that the proliferation of disease is more a result of physiological stress due to a combination of effects, rather than the existence of bacteria, alone (Shinn et al. 2000; Lesser et al. 2007; Rypien 2008; Raymundo, Couch, and Harvell 2008).

Of the diseases monitored in the Caribbean, four are suspected to have a connection with Africa dust events: aspergillosis, black band disease, white pox, and white plague (Garrison et al. 2003). Aspergillosis, black band disease is well-characterized. Experiments suggest that the relationship between black band disease and African dust transport depends largely on iron concentration of the dust. In fact, it is suggested that the iron facilitates the pathogenicity of the black band cyanobacterium (Garrison et al. 2003). Two types of white plague have also affected Caribbean coral, even in the most pristine regions (Garrison et al. 2003). The pathogen of one type of white plague was identified as the bacteria *Aurantimonas coralicida* (Denner et al. 2003), which is also suspected to cause the most damage to coral reefs in the Caribbean in the presence of iron-rich dust (Garrison et al. 2003). Finally, white pox does not have a definitive origin in African dust but is hypothesized to be introduced to coral reefs via several possible sources, including soil runoff, river discharge, and atmospheric transport of soil (Garrison et al. 2003).

2.3. Remote sensing of coral reefs and the environment

Disease proliferation in the Caribbean Sea caused by dust-borne pathogens cannot be directly observed using remote sensing. Environmental variables that affect coral reefs, such as SST, are already monitored using satellite remote sensing, and the effect of pathogens on coral reefs can be inferred from the correlation of these data sets. These variables can be used as proxies to determine relationships between coral reefs, disease, and the environment.

As described in Mumby et al. (2004), aerosols such as African dust have both a positive and negative impact on coral reefs by altering Earth's radiative balance, thereby controlling the amount of thermal stress experienced by corals, and by being deposited on the ocean surface. Because dust aerosols absorb most strongly in the UV region of the electromagnetic spectrum, certain sensors can distinguish the dust from other background aerosols such as pollution, which reflect UV radiation. The total ozone mapping spectrometer (TOMS) and its successor, the ozone monitoring instrument (OMI), both measure the solar irradiance and the backscattered radiance from Earth in this region to determine aerosol optical thickness and derive values for AAI.

The light field of ocean surface waters is described by the diffuse light attenuation coefficient, k . This coefficient depends on the wavelength of incident radiation and is a measure of the clarity of the water, which is a function of the concentration of dissolved organic material and suspended sediments (Mumby et al. 2004). The combination of estimated water column attenuation coefficients, UV, and optical data are used to estimate the amount of solar radiation received by marine organisms. This parameter is currently measured by several sensors, including MODIS, the Sea-viewing Wide Field-of-view Sensor (SeaWiFS), the medium resolution imaging spectrometer, and the ocean colour monitor.

Incident solar radiation at the ocean surface drives SST variations. Long-term measurements of SST collected by the advanced very high-resolution radiometer (AVHRR) are used by the national oceanic and atmospheric administration (NOAA) coral reef watch programme to derive a suite of satellite data products to quantify coral bleaching-related heat stress (Mumby et al. 2004; Eakin et al. 2010). The data products are derived from the AVHRR-produced SST climatology at a 50 km spatial resolution and were used in 2005 by NOAA to warn scientists of developing positive temperature anomalies over the Caribbean region and of the subsequent coral bleaching and mortality that followed (Eakin et al. 2010).

Although iron is one of the most important constituents of African dust, the use of remote sensing to study its direct effects on coral reefs is limited (Gasso, Grassian, and Miller 2010). This limitation is due to the need for additional information regarding the state of the mixed layer and concentration of other nutrients, which modulate the response of phytoplankton to dust (Gasso, Grassian, and Miller 2010). Regions characterized by high iron concentration, however, are also characterized by high concentrations of chl- a , which is measured by satellite ocean colour sensors (Mahowald et al. 2005). Satellite ocean colour measurements by instruments such as SeaWiFS are used to estimate the concentration of suspended material near the ocean surface. Andréfouët et al. (2002) suggested that ocean colour observations can be used to infer and track the primary pathways for dust and pathogens within and among coral reef regions in order to assess the response of coral reefs to short-term events. Erickson III et al. (2003) used SeaWiFS ocean colour measurements to uncover correlations between atmospheric iron and chlorophyll concentrations on the ocean surface in the Patagonian region. This research concluded the existence of a strong temporal correlation between ocean surface monthly averages of simulated atmospheric iron deposition and chlorophyll measured by SeaWiFS (Erickson III et al. 2003).

2.4. Data mining

NASA's Earth observation programme provides continuous observations of the entire Earth and generates an unprecedented amount of science data. These data are heterogeneous, dynamic in space and time, contain multivariate connections, and contain both explicit and implicit spatial relationships and interactions (Mennis and Guo 2009). Advanced data mining techniques are required to capture the spatio-temporal relationship, autocorrelation, and idiosyncrasies of the data. Traditional statistical methods focus primarily on univariate spatial autocorrelation and are unable to process large volumes

of data. Further, traditional methods are not efficient at discovering embedded patterns and information in vast data sets (Mennis and Guo 2009).

In general, data mining involves the analysis, recognition, and establishment of associations and patterns in data (Lausch, Schmidt, and Tischendorf 2015). The mining of frequent item-sets in a data set uncovers associations, correlations, and causal relationships among items (Han, Kamber, and Pei 2011). Similarly, spatio-temporal data mining involves the extraction of knowledge and the analysis of vast data sets to find associations, structures, patterns, and other irregularities not explicitly stored in spatio-temporal data (Lausch, Schmidt, and Tischendorf 2015).

The standard approach to find relationships among ecological events across different parts of the globe requires the computation of correlation between spatio-temporal time series, and then identifying the regions with highest correlation (Tan et al. 2001). This approach is the basis for ARDM, which was initially developed for market analysis (Tan et al. 2001).

Association rules are expressed in notation as $\{A, B\} \rightarrow C$, where A and B are called the antecedents and C the consequent. An antecedent is comprised a set of items, known as an item-set, or a group of item-sets. Each rule describes the likelihood that C will occur when the item-set occurs. The total number of antecedents in an item-set can vary between one and the total number, N , of unique items in the data set, and the number of consequents can vary on a similar range. This study uses a total of five antecedents, where each antecedent represents one of the environmental parameters described in Section 2. This study uses the total number of coral disease observations during each sampling period as the single consequent.

When developing association rules from large transaction databases, each rule is evaluated by two measures: confidence and support. The confidence of a rule can be expressed mathematically as follows

$$\text{Conf}(\{A, B\} \rightarrow C) = P(C|\{A, B\}) \quad (1)$$

In other words, this measure describes the conditional probability of C given A and provides the accuracy of the rule. The support of a rule describes the probability of transactions containing both antecedent and consequent and provides awareness of how often the rule is relevant. Support is calculated mathematically as in Equation (2)

$$\text{Sup}(\{A, B\} \rightarrow C) = P(\{A, B, C\}) \quad (2)$$

While support and confidence provide measures of whether or not an association rule is strong, it is possible for rules with high confidence to have the lowest correlation, or the rules with the lowest confidence to have the highest correlation. This is because, as noted in (Tan et al. 2001), the calculation of confidence is agnostic to the calculation of support. Therefore, a correlation measure of interestingness known as lift is typically used to rank association rules. Lift expresses the correlation between the antecedent and the consequent and is written mathematically as in Equation (3)

$$\text{Lift}(\{A, B\} \rightarrow C) = \frac{P(\{A, B\} \cup C)}{P(\{A, B\}) \times P(C)} \quad (3)$$

Lift values less than one represent a negative correlation between the antecedent and consequent; in other words, there is a probability that the occurrence of one of these values, or sets of values, leads to the absence of the other value (Han, Kamber, and Pei

2011). On the other hand, lift values greater than one represent a positive correlation, in which the occurrence of a value or set of values infers the occurrence of the other value (Han, Kamber, and Pei 2011). Should the lift value equal one, the antecedents and consequent are independent and no correlation exists.

Mennis and Liu (2005) demonstrated one successful application of ARDM to spatio-temporal data. In their study, they determined associations among processes of socio-economic change and urban growth using data from the US Census Bureau and land-cover data from USGS aerial photography. Similarly, Rajasekar and Weng (2009) applied ARDM to a combination of data from the advanced spaceborne thermal emission and reflection radiometer, scaled normalized difference vegetation index, population density data, and land-use land-cover data to explore relationships between land-surface temperature and urban change. This study demonstrated the utility of ARDM to extract quantitative information regarding the relationships among urban landscape parameters derived from multiple data sets.

3. Methods

3.1. Study area

This study covers the time period between 1 August 1999 and 31 December 2010, and the region of study is the Caribbean Sea, shown in Figure 1, with bounds 9.5°N, 24.5°N, 87.5°W, and 60.5°W.

3.2. Methodology

3.2.1. Data acquisition

Satellite data from four sensors were used in this study, as shown in Table 1. UV-AAI represents the back-scattered radiance in the UV and indicates the presence of desert dust in the atmosphere over the ocean and land. Daily level 3 UV AAI data from the



Figure 1. The study area. Every coral disease observation from 1999 to 2010 is represented by a red asterisk (<http://www.reefbase.org>).

Table 1. Satellite data used for this study.

Satellite sensor	Product	Spatial resolution (°)	Temporal resolution	Dates collected	Parameter	Format
EP-TOMS	L3 UV-AAI	1 × 1.25	Daily	1 August 1999–31 December 2004	AAI	ASCII
Aura OMI	OMAEROe (L3 UV-AAI)	0.25 × 0.25	Daily	1 January 2005–31 December 2010	AAI	HDF-5
OrbView-2 SeaWiFS	L3 SMI K490	5 × 5	8 day average	1 August 1999–31 December 2010	Diffuse attenuation at 490 nm	HDF- EOS
OrbView-2 SeaWiFS	Chl- <i>a</i>	5 × 5	8 day average	1 August 1999–31 December 2010	Chlorophyll concentration	HDF- EOS

Earth probe (EP) TOMS and the Aura OMI were obtained from the Goddard Earth Science (GES) Data and Information Services Center (DISC). In both UV-AAI data sets, a value of 0 implies the existence of clouds, while values larger than 0 imply the presence of UV-absorbing aerosols.

The diffuse light attenuation coefficient at 490 nm (K490) is a qualitative description of water clarity in the blue-green portion of the electromagnetic spectrum and is influenced by the concentration of dissolved organic materials (Mumby et al. 2004). Sunlight incident on the ocean surface can either be absorbed or scattered, depending on the amount of phytoplankton, suspended sediment, and dissolved organic material that is present in the water column (Capone and Subramaniam 2005). Ocean colour sensors such as SeaWiFS measure the back-scattered radiance in the visible wavelength band to characterize the magnitude and variability of chlorophyll production on the ocean surface. This information also suggests the role of the ocean in certain biogeochemical cycles, such as that of iron. Chl-*a* measurements, therefore, can be used as a proxy to observe nutrient density on the ocean surface. Level 3 standard mapped images of 8 days composite K490 and chl-*a* measurements were obtained from the Ocean Biology Processing Group (NASA Goddard Space Flight Center, Ocean Ecology Laboratory, Ocean Biology Processing Group 2014).

Non-satellite data are described in Table 2. SST is the temperature of the top millimetre of the ocean's surface and is measured by several instruments including the legacy AVHRR. SST data were obtained from NOAA's Earth System Research Laboratory (ESRL) physical sciences division. The optimum interpolation SST data product, based on data from the Pathfinder AVHRR reanalysis, aqua advanced microwave scanning radiometer, and spatially smoothed, 7 days average *in-situ* buoy measurements, provides continuous, unbiased SST data from 1985 to the present (NOAA Earth System Research Laboratory. n.d.).

Table 2. Non-satellite data used in this study.

Product	Source	Spatial resolution (°)	Temporal resolution	Dates collected	Parameter	Format
OI-SST (v. 2.0)	NOAA	1 × 1	Monthly	1 August 1999–31 December 2010	SST	NetCDF
SOI	NOAA	N/A	Monthly	1 August 1999–31 December 2010	El Niño	TXT
UNEP GCDD	ReefBase	N/A	Yearly	1 August 1999–31 December 2010	Disease Observations	MS Excel

The SOI is a single value for each month that indicates the existence and intensity of El Niño or La Niña. It is defined as the normalized difference in measured sea-level pressures at Darwin, Australia, and Tahiti. A positive value of SOI indicates the existence of La Niña, while a negative value of SOI indicates the existence of El Niño. To account for the influence of ENSO on coral disease, SOI values data were obtained from the NOAA ESRL physical sciences division archive of climate indices.

Observations of coral disease were obtained from Reefbase GIS (Reefbase 2016), an online database originally compiled by the UN Environmental Programme (UNEP) World Conservation Monitoring Centre (WCMC) in conjunction with the NOAA National Marine Fisheries Service (NMFS). The original database was called the Global Coral Disease Database (GCDD) (United Nations 2016). The GCDD was comprised of data mined from peer-reviewed literature, technical reports, and volunteered data from individual and organizational fieldwork. Total observations of various coral disease types thought to be related to African dust are shown in Table 3.

Prior to 2016, the database was updated continually by users and provided a central platform for aggregation and dissemination of disease data in a consistent format. Although the data set was useful, it has a coarse temporal resolution of 1 year. The original observations submitted to UNEP-WCMC by users included the date, month, and year of each observation; however, it was not possible to obtain metadata for a large number of observations directly from the GCDD, at the time this study was conducted. The ReefBase data set, on the other hand, provides only the year of each observation. In spite of this limitation, this study uses the ReefBase data product because it provides the best available data over a large spatial scale and long time period. Due to the low sampling rate compared to that of the satellite data, a piece-wise cubic Hermite interpolating polynomial was applied to interpolate disease counts between years and obtain estimates every 3 months. This method avoids increasing the temporal resolution

Table 3. Total counts of all coral disease types or anomalies observed in the study region from 1999 to 2010, including predation.

Disease type	Count	Location(s)
Aspergillosis	20	Colombia, Belize, Bahamas, Dominican Republic, Honduras
Black blotch	2	Jamaica
Bacterial bleaching	47	Colombia, Mexico
Black-band disease	131	St. Lucia, Costa Rica, Venezuela, Colombia, Netherlands Antilles, St. Vincent, Puerto Rico, Mexico, US Virgin Islands, Cayman Islands, Cuba, Turks and Caicos, Jamaica, Cuba, Panama
Dark-spot disease	113	Colombia, Venezuela, Netherlands Antilles, US Virgin Islands, Puerto Rico, Cayman Islands, Cuba, St. Lucia, Jamaica, Mexico, Trinidad and Tobago
Unspecified	387	Costa Rica, Venezuela, Netherlands Antilles, St. Vincent, US Virgin Islands, Mexico, Cayman Islands, Turks and Caicos, Colombia, Jamaica, Cuba, Panama, St. Lucia, Honduras, Haiti, Puerto Rico, Bahamas, Dominican Republic, Dominican Republic, Guadeloupe, Montserrat, British Virgin Islands, Aruba, Martinique, Trinidad and Tobago
Fungal disease	6	Mexico
Parrot fish predation	2	Mexico
White pox	23	Colombia, Cuba, Mexico, Puerto Rico
Red-band disease	5	Turks and Caicos, Costa Rica, Jamaica, Netherlands Antilles, Mexico
White plague	462	Colombia, US Virgin Islands, Puerto Rico, Turks and Caicos, Cuba, Cayman Islands, Mexico, St. Vincent, Netherlands Antilles, Costa Rica, Venezuela, Jamaica, Panama, Dominican Republic, Trinidad and Tobago

of the satellite data products, since the use of annual averages in these data would result in loss of information. This method inherently assumes that the distribution of disease observations is consistent with the actual trend of coral disease over time.

3.2.2. Preprocessing and preparation

Due to differences in spatial resolution between satellite sensors, the data used in this study were re-gridded to a common $2^\circ \times 2^\circ$ grid in MATLAB by applying a Delaunay Triangulation algorithm and standard linear interpolation on the resulting data space. After re-gridding, each data set was processed to create three-dimensional 'stacks' of data with time along the Z-axis, longitude along the X-axis, and latitude along the Y-axis.

The spatio-temporal nature of satellite-based Earth science data presents several challenges for traditional data mining techniques. In particular, the temporal autocorrelation of data can have an impact of the significance of any computed statistical correlation (Tan et al. 2001). Reducing the short-term autocorrelation in time-series data can be accomplished by aggregating the data into bins of 3 months intervals (Tan et al. 2001). Subsequently, each time series, including the coral disease time series, was reprocessed into bins containing average values over a period of 3 months. This average inherently positioned the resulting data point in the middle of the 3 months period.

Finally, seasonal variations in time series data are usually removed because the seasonal cycle dominates the original signal, hindering the detection of anomalies (Tan et al. 2001). In Earth science, in particular, it is advantageous to relate some event in some location to an anomalous climate condition occurring in the same location, region, or completely different part of the world. Anomalous climate conditions are only apparent if periodic fluctuations corresponding to the seasons are removed. Additionally, the seasonal signal is generally removed to make the original time series stationary. The seasonal signal was removed from each time series using a 12 months moving average. The primary limitation of the 12 months average, although not complex to implement, is its tendency to spread the effects of deviations from average values across neighbouring points in time (Tan et al. 2001).

For the ARDM algorithm, each data set was transformed into a binary table of transactions in which each row of the table represents a set of events. A binary transaction table describes the occurrence, or lacks thereof, of events and expedites the generation of patterns that show the co-occurrence of these events. To create the table, each data set was discretized into two clusters of similar characteristics. ARDM, having been developed for market analysis, is designed to operate with nominal and ordinal data, not numeric data (Mennis and Liu 2005). Since spatio-temporal data are numeric, the data must be discretized into specific categories, i.e. low or high, near or far. This discretization process removes some intrinsic information and can have a large impact on ARDM results. Nevertheless, discretization is an essential procedure to the ARDM algorithm because the number of discrete values in a geophysical time series can be very large, resulting in an inordinate number of association rules with statistically low significance. Selection of the discretization algorithm should consider the distribution of the data set and the closeness of the data values. Since ARDM was originally designed to deal with Boolean-like values, each data set is divided into low and high values based on some predetermined threshold value (Yang, Tang, and Sun 2011).

In this analysis, two discretization methods were used (1) in the simplest method, the mean value of each data set is used the threshold to partition high and low values, and (2) in the more complex method, a *K*-means clustering function is used. The *K*-means function implementation partitions points in a two-dimensional matrix into *k* number of clusters by finding *k* centroids and assigning each point to the cluster associated with the nearest centroid. The process of selecting centroids seeks to minimize the total error for each point, where the error is defined as the discrepancy between a point and its centroid, or the squared distance (Steinbach, Karypis, and Kumar 2000). The input to the *K*-means function is a data vector and the number of clusters, which was set at a value of two. The resulting output was a one-dimensional vector containing the cluster indices of each point. This vector was input into custom code to match each data point to the appropriate cluster. The resulting bin centres are given in Table 4.

Data from the Reefbase coral disease database spanned from 1999 to 2010, but, as the ReefBase data set contained no points for 2006 within the study region, this year was not included. Each entry in the coral disease database was binned into the appropriate latitude and longitude on a $2^\circ \times 2^\circ$ grid. After binning each data entry, a time series of disease observations was generated at each grid location. Each time series was then interpolated using a piecewise cubic Hermite interpolating polynomial function due to the sparse temporal resolution of the original data set. To visualize the data, consider the region off the southern coast of Jamaica, as shown in Figure 2. This region covers 15.5° – 17.5° N and 78.5° – 76.5° W. The original time series, describing area-averaged coral disease observations over the study period, is shown on the top *X* and right *Y* axes in Figure 3, along with the time series after interpolation along the bottom *X* and left *Y* axes, where the *X*-axis represents 3 months' bins. The centroid of each 3 months bin is the 2nd month.

The distribution of the disease data set following discretization was biased towards low values due to its spatial and temporal sparseness. The mean of this distribution was unsuitable as a threshold for discretization because the threshold has the potential to be set very high by a small number of large disease counts. Instead, a piece-wise function was used on values below threshold. This threshold was set to 0.3 because the sequence of interpolating the 12 months data set and binning the resulting 1 month data set into bins containing 3 months' averages generated values less than 1. After using this threshold to delineate two low and high bins, those values at the extreme end of the data set, greater than one, were binned into the high category. All 0 values were discarded from this data set prior to analysis. The GCDD does not contain continuous observations at each latitude and longitude over the study period; therefore, values of 0 in each time series generated for this study do not necessarily correspond to observations of non-diseased coral and may simply correspond to missing data. The resulting

Table 4. Centre points of bins for each satellite data set after applying the discretization algorithms.

Parameter	Low	High
SST ($^\circ$)	24.25	28.15
Chl- <i>a</i> (mg m^{-3})	0.23	8.09
K490 (m^{-1})	0.042	0.897
UV-AAI (unitless)	≤ 0	≥ 0

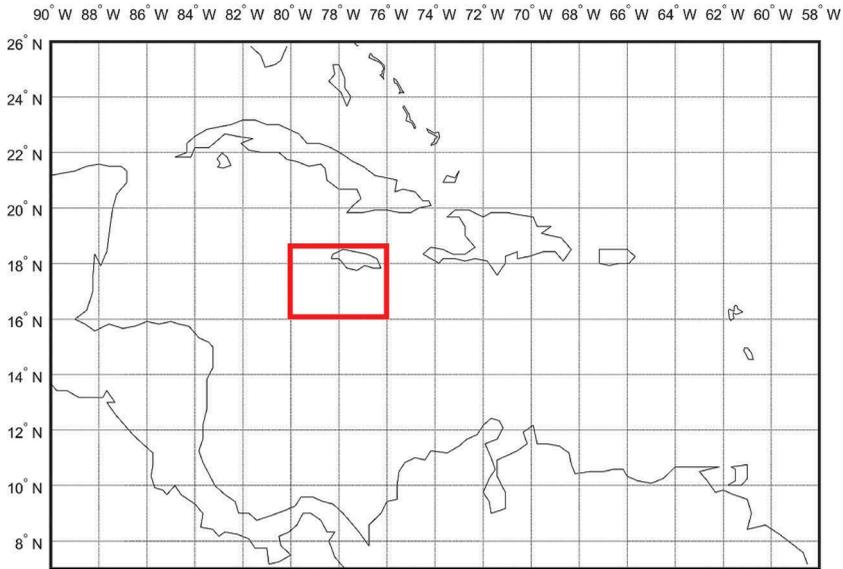


Figure 2. Region (in red) for which the time series in Figure 3 was extracted and plotted.

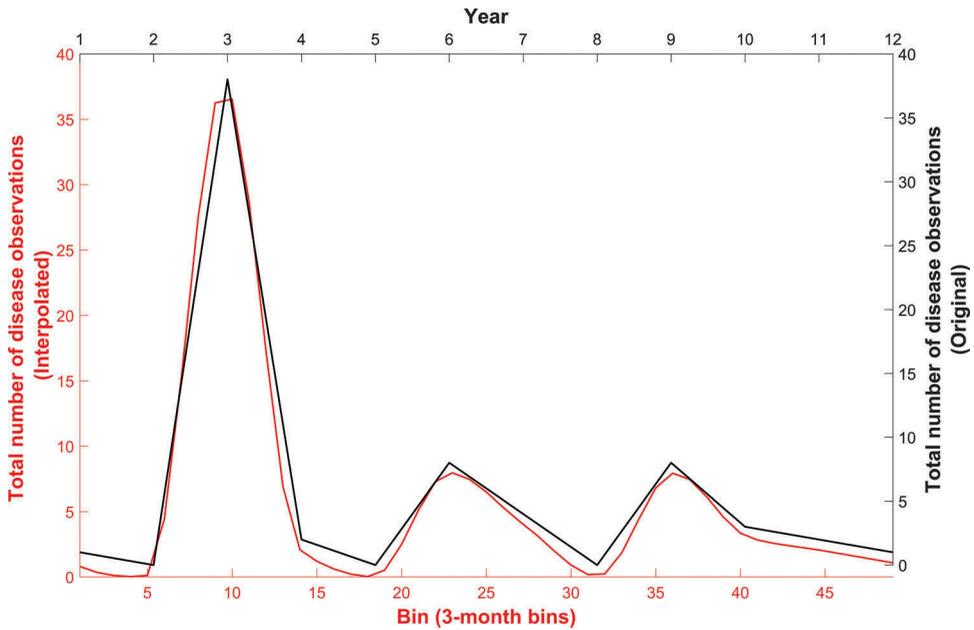


Figure 3. Time series of original disease counts from the GCDD (black) and interpolated disease counts (red).

discretized values of coral disease observations, therefore, represent different degrees of available observations, not the absence or presence of observations.

As with the disease data, the AAI data required a different discretization routine than the other remote-sensing data. Figure 4 shows the distribution of AAI before discretization. AAI ranges are defined such that values less than 0 represent non-UV-absorbing

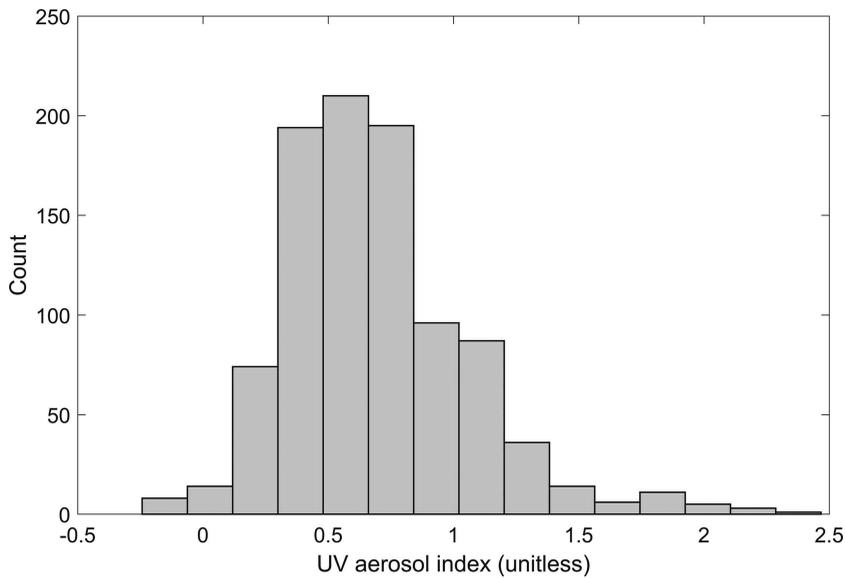


Figure 4. The distribution of UV AAI values over the entire region and study period, before discretization.

aerosols, while values greater than 0 represent UV-absorbing aerosols, such as dust, and values close to 0, within a variance of 0.1, represent clouds. These rules for AAI were coded into MATLAB as a piece-wise function and discretization was performed on the AAI data set using this function. The resulting distribution of AAI values following discretization are shown in Figure 5.

The *K*-means clustering algorithm was used to discretize the SST, chl-*a*, and K490 data sets. When applied to SST, the algorithm created two clusters, where the cluster classified as low was centred at 27.25°C, and the cluster classified as high was centred at 28.15°C. When applied to chl-*a*, the *K*-means algorithm classified low values with a cluster centred at 0.25 mg m⁻³ and high values with a cluster centred at 8.09 mg m⁻³. Research shows that the average chl-*a* concentration measurement in the waters around Cuba is 0.07 mg m⁻³ (Gonzalez et al. 2000); therefore, the histograms imply that for the entire study period and study region, chl-*a*s did not deviate significantly from the average and were, in fact, significantly lower. To verify that the average value of chl-*a* remained near average over the entire study period, the web-based GIOVANNI Interactive Visualization and Analysis Tool available from GES-DISC was used to visualize the average Sea-WiFS-measured chl-*a* between September 1999 and December 2010. The average concentration is shown in Figure 6. Similarly, the distribution of K490 in the region over the entire study period was also heavily weighted towards low values. The *K*-means clustering algorithm created a low cluster centred at 0.04 m⁻¹ and a high cluster centred at 0.90 m⁻¹. In literature, a K490 value of 1.0 m⁻¹ is considered 'high'.

3.2.3. Association Rule Data Mining (ARDM)

Preprocessed remote-sensing and coral disease observation data used for input to ARMADA (Malone 2011) were first saved to comma-separated value files, where each

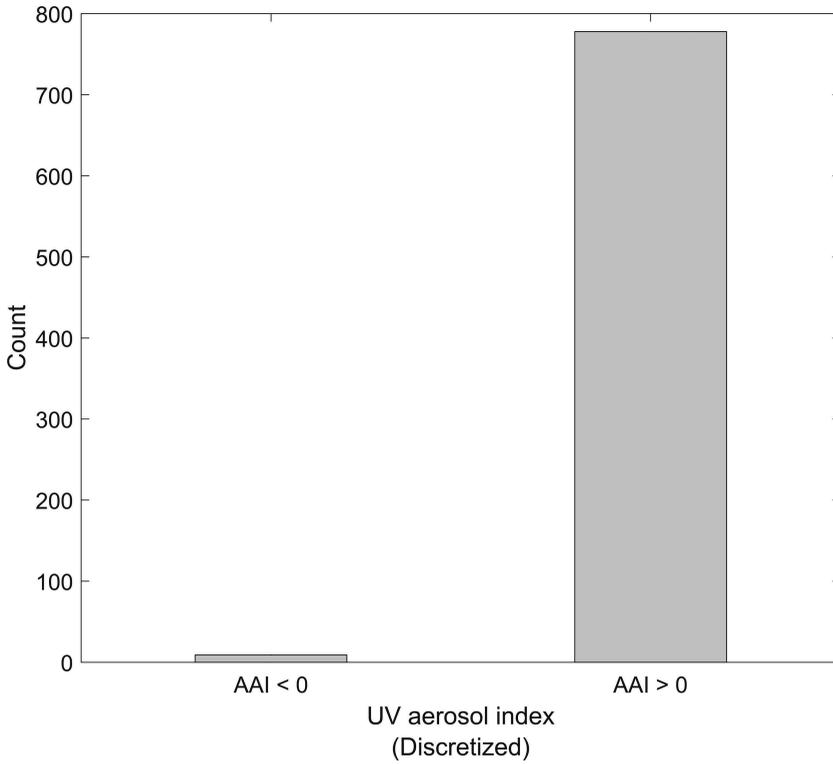


Figure 5. The distribution of UV AAI values after discretization.

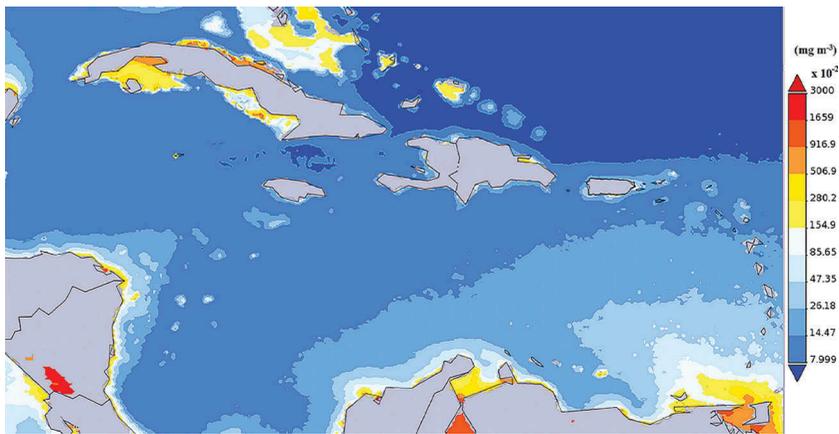


Figure 6. Time-averaged chlorophyll-*a* concentration for the study period.

antecedent and consequent for each transaction were listed on a single row, separated by commas. An example of an input transaction table, prior to discretization, is in [Table 5](#). To maximize the number of association rules generated for each experiment, the support and confidence criteria were set to their minimum values of 1%.

Table 5. Example input transaction table for ARDM analysis prior to discretization.

Time (month of year)	Lat. Bin	Lon. Bin	SST (°)	Chl- <i>a</i> (mg m ⁻³)	K490 (m ⁻¹)	UV-AAI	Disease	SOI
1	1	1	27.04	0.19	0.046	NaN	0	-4.4
2	1	1	26.20	0.32	0.057	NaN	0	-3.4
3	1	1	27.65	0.91	0.116	NaN	0	-4.0
4	1	1	27.32	0.20	0.045	NaN	0	-2.4
5	1	1	26.38	0.27	0.052	NaN	0	0.4
6	1	1	25.27	0.36	0.062	NaN	0	1.6
7	1	1	28.10	0.78	0.092	NaN	0	2.0
8	1	1	27.77	0.36	0.062	NaN	0	1.9
9	1	1	27.37	0.21	0.050	NaN	0	1.7
10	1	1	26.07	0.33	0.061	NaN	0	1.8
11	1	1	28.14	0.85	0.097	NaN	0	1.7
12	1	1	27.90	2.12	0.214	NaN	0	2.3

Additionally, mining goals were set to extract only those goals where the consequent was equal to a value in the coral disease observation time series.

3.2.4. Significance testing

ARDM explores large data sets to find patterns that satisfy some type of constraint. Traditionally, rules are constrained by minimum support and minimum confidence values, where confidence is considered to measure a rule's strength and support is considered to measure a rule's statistical significance (Webb 2007). One serious limitation of the rules developed by these techniques, however, is that there is a large risk of generating spurious rules. Overcoming this problem requires evaluating the generated rules using statistical significance tests. Statistical significance tests address the question of whether or not the computed relationships between the given antecedents and consequent actually exist. To calculate the statistical significance of rules generated by ARDM, Webb (2007) suggested the use of the Fisher's exact test. This test is used to determine whether non-random associations exist between two variables by calculating the hypergeometric probability of observing a given set of values. The null hypothesis is that the given association rule or antecedent does not satisfy the minimum support and minimum confidence constraints specified for the algorithm; that is, that a given rule is a false discovery.

The Fisher's exact test was applied to each parameter using the R statistical language. The fisher test function in the R stats package examined whether or not a given parameter was significant with regard to its relationship to coral disease observations. The function generated five 2 × 2 contingency tables containing the support of the relationship expressed by the corresponding column and row, as shown in Table 6.

Table 6. Contingency table for SST and disease observations.

	Number disease observations (L)	Number disease observations (H)
SST (L)	163	55
SST (H)	169	208

L represents a low number of disease observations or low value of SST, and H represents a high number of disease observations or high value of SST. Ranges are defined by the discretization algorithm as described Section 3.2.2.

These values represent the probability of that parameter occurring with the consequent, without regard to other parameters, and are calculated by the ARDM algorithm.

4. Results

The ARMADA tool generated a total of 146 rules, ranging from rules with one-item antecedents to those with five-item antecedents, in which all parameters were used in the analysis. Redundant rules, those containing fewer parameters than the total number of parameters, are a subset of the rules containing all parameters. Redundant rules are removed manually by comparing each rule with antecedent sets containing less than five items to those containing exactly five items. Results are then further pruned by calculating lift values for each rule and selecting only those rules with lift values greater than one. Select association rules generated for the entire study region over the 1 August 1999 to 31 December 2010 time period are shown in Table 7. *H* represents the high state of a parameter, and *L* represents the low state of a parameter. Since SOI is a single value for every month, a positive value of SOI indicates the existence of La Niña, and a negative value of SOI indicates the existence of El Niño.

The first rule conveys that a greater number of disease observations is probable when SST is high, chl-*a* concentrations are low, K490 values are low, dust is present, and El Niño is dominant. The lift calculated for this rule is 1.5366, which means that the elements of this rule can be described as having a positive correlation, that is, the occurrence of this particular combination of geophysical states infers the occurrence of a greater number of disease observations. This rule would be considered *strong* because its lift value is much greater than one. The second rule is similar; however, it shows that the same state values for SST, chl-*a*, K490, and dust are likely to occur with a greater number of disease observations even when La Nina is present. The lift value calculated for this rule is 1.0947, indicating that although the correlation is considered positive, it is not as strong of a correlation as that of the first rule, which favoured the presence of El Niño. The third rule shows that a lower number of disease observations are likely when SST, chl-*a*, and K490 values are low, dust is present, and La Nina is the dominant pattern. Here, the lift value is 1.0742, implying the correlation is again positive, but not strong, since it does not deviate much from a value of one. The fourth and fifth rules do not contain all parameters; however, they are included in the table because they represent non-redundant, unique rules with high lift values. The fourth rule shows that a lower number of disease observations are likely when chl-*a* concentrations are low, K490 values are low, and dust is not present. The lift value calculated for this rule is greater than one and greater than that of the previous rule, implying a greater likelihood of a low number of disease observations co-occurring with the presence of non-UV-

Table 7. Select association rules generated for the study region over the 1999–2010 study period.

Rule	SST	Chl- <i>a</i>	K490	UV-AAI	SOI	Disease	Support	Confidence (%)	Lift
1	H	L	L	H	L	H	104	37.8	1.5366
2	H	L	L	H	H	H	104	26.9	1.0947
3	L	L	L	H	H	L	92	26.4	1.0742
4	L	L	L	L		L	15	48.4	1.9661
5	H	L			L	L	71	25.5	1.0340

Table 8. *P* values for each environmental parameter discussed in this article.

Parameter	<i>p</i> Value
SST	<<0.0001
AAI	0.0001
SOI	0.0208
K490	1.0000
Chl- <i>a</i>	1.0000

absorbing aerosols. The final rule shows that a lower number of disease observations are likely even when SST is in a high state, if it occurs with low chl-*a* concentrations, low K490 values, and an El Niño pattern.

P values calculated and compared against a standard 95% confidence level for Fisher's exact test are presented in Table 8. This table conveys that given the distributions of the original data sets, the least statistically significant parameters with regard to coral disease observations are K490 and chl, with *p* values of 1, while the most statistically significant parameters are AAI, SOI, and SST, with *p* values of 0.0208, 0.0001, and much less than 0.0001, respectively. The results of Fisher's exact test provide qualitative measures to describe the relative significance of each parameter with regard to the number of coral disease observations and suggest that SST is the most significant driver of such observations.

5. Discussion

5.1. Impact of the ARDM algorithm

The rules generated using the ARDM algorithm indicate a relationship between African dust and coral disease observations in the Caribbean Sea, modulated by SST. While the result that coral disease is driven primarily by SST is not a new finding, the results of this study support the hypothesis that African dust plays a role, as well, in conjunction with other environmental parameters.

Fisher's exact test produces a *p* value much less than 0.0001 with regard to the relationship between SST and coral disease observations. For values of $p < 0.05$, the null hypothesis is rejected, as it is for SST. This indicates that SST is a statistically significant parameter in determining the prevalence of coral disease. This finding supports research such as that of Selig et al. (2006), which hypothesizes that rising ocean temperatures may be exacerbating the effects of infectious disease on coral reefs.

The first two association rules shown in Table 7 indicate that a greater number of disease observations are positively correlated to the presence El Niño with a high SST and the presence of dust, while less positively correlated to the presence of La Niña. This non-definitive result suggests a weak correlation between disease observations and the state of the El Niño, and the *p* values calculated using Fisher's exact test support this hypothesis. The *p* value for SOI is 0.0208 and because it is close to 0.05, it is considered only marginally statistically significant. Previous research, however, has identified correlations between El Niño and African dust. Specifically, the annual variation of exported dust from the Sahara is shown to be driven by El Niño by Garrison et al. (2003), Griffin and Kellogg (2004), and Goudie and Middleton (2001). These studies show that the

Sahara experiences drier than normal climate conditions in the presence of a strong El Niño, contributing to high levels of dust concentration transported to the Caribbean Sea.

5.2. Statistical significance

Our results consider SST a statistically significant parameter with regard to coral disease, and previous research shows a correlation between El Niño and African dust. Additionally, the first rule in Table 7 indicates that disease observations are highly correlated to the combined presence of dust, high SST, and El Niño; therefore, we infer that this rule supports the hypothesis that African dust storms influence the prevalence of coral disease observations.

The results of Fisher's exact test also suggest that SST influences the prevalence of coral disease in the Caribbean more than other environmental parameters considered. The p values calculated for both chl- a and K490 are one, inferring that these parameters are not statistically significant with regard to coral disease observations. As described in Section 3, these results should be used with caution. The chl- a and K490 data sets used for this study are biased towards low values. To mitigate this bias, further research should consider a greater temporal range.

These results do not explicitly describe a relationship between dust concentrations and SST but do infer a tightly coupled relationship between the prevalence of coral disease and SST in the presence of dust. Given the results of the data mining algorithm and the significance test, this study supports the hypothesis that the influence of African dust storms on the number of coral disease observations is primarily dependent on ocean temperature, rather than on turbidity and nutrient concentration.

5.3. Visual analysis

To understand the results from the application of the data mining algorithm as well as those of Fisher's exact test, data for the study region were obtained for *quiet* and *active* years of dust storms using the web-based GIOVANNI Interactive Visualization and Analysis Tool available from GES-DISC. To distinguish between quiet and active years, a time series of UV AAI values from TOMS-EP were plotted between 2000 and 2005, without removing the seasonal signal, as shown in Figure 7. From this plot, 2001 was selected to represent a quiet year due to the dominance of AAI values below 1, implying a relatively low concentration of UV-absorbing aerosols as compared to other years. The year 2003 was then selected to represent an active year because the maximum AAI value on this time series was found to occur between 2003 and 2004.

For each year in the analysis, the total number of coral disease observations was obtained from the GCDD. In 2001, a total of 173 counts of coral disease were observed in this region, while a total of 221 counts were observed in 2003. Figure 8 shows the transport of African dust in 2001, represented by AAI, during the peak season for dust transport. This image was generated using GIOVANNI and data from TOMS-EP.

For the entire year 2001, time series of area-averaged SST, chl- a , and K490 was generated using GIOVANNI and Sea-WiFS data. Shown in Figure 9, the average trend for chl- a in this region did not deviate significantly from an average of 0.12 mg m^{-3} . Similarly, as shown in Figure 10, K490 was not found to deviate significantly from an

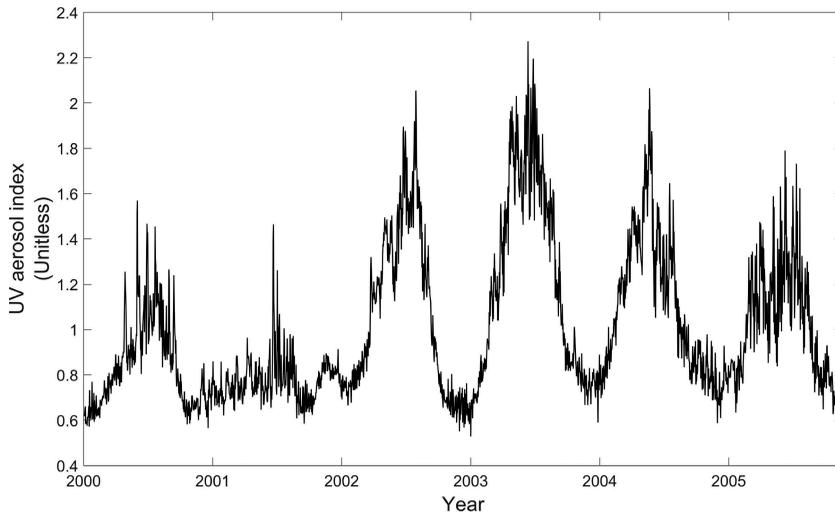


Figure 7. Area-averaged daily TOMS-EP UV AAI between 2000 and 2005.



Figure 8. Time-averaged map of UV AAI measured by TOMS-EP for June 2001, depicting UV-absorbing aerosols in the atmosphere over the Saharan region of Africa.

average value of 0.03 m^{-1} . **Figure 11** shows a monthly SST as measured by MODIS Terra – because MODIS Aqua was not available until 2002 – and it shows a trend centred on an average of 27.5°C and a maximum value at approximately 29.5°C .

The same process described previously was performed on data from 2003. **Figure 12** shows the transport of African dust in 2003, represented by AAI, during the peak season for dust transport. As for the previous images, this image was generated using GIOVANNI and data from TOMS-EP. These images show that 2003 was dramatically more active in terms of dust storms than 2001.

Figure 13 shows that the average trend for chl-*a* in this analysis region deviated significantly from the regional average of 0.07 mg m^{-3} to an average of approximately 0.15 mg m^{-3} , with a maximum value of approximately 0.16 mg m^{-3} during the period of maximum African dust extent. Similarly, as shown in **Figure 14**, K490 was found to trend higher than in 2001, with an average of approximately 0.04 m^{-1} . **Figure 15** shows a

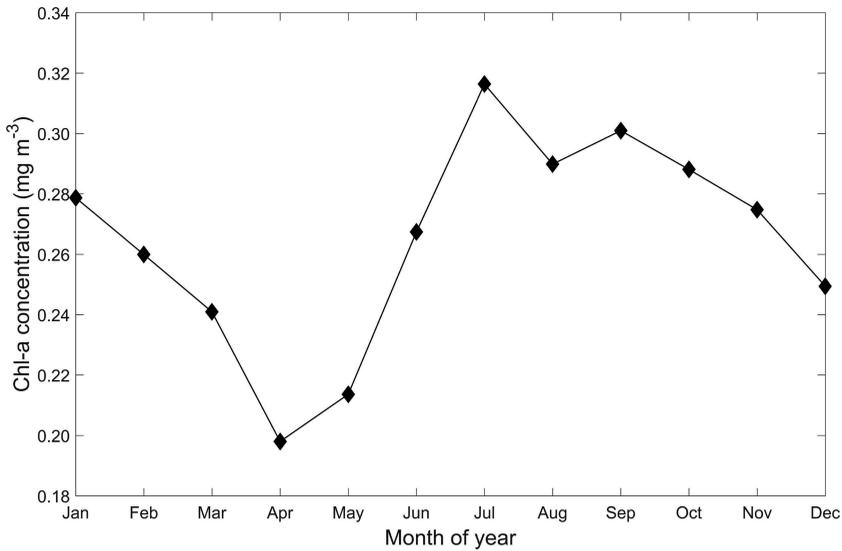


Figure 9. Sea-WiFS area-averaged time series of chl-*a* concentration between January 2001 and December 2001.

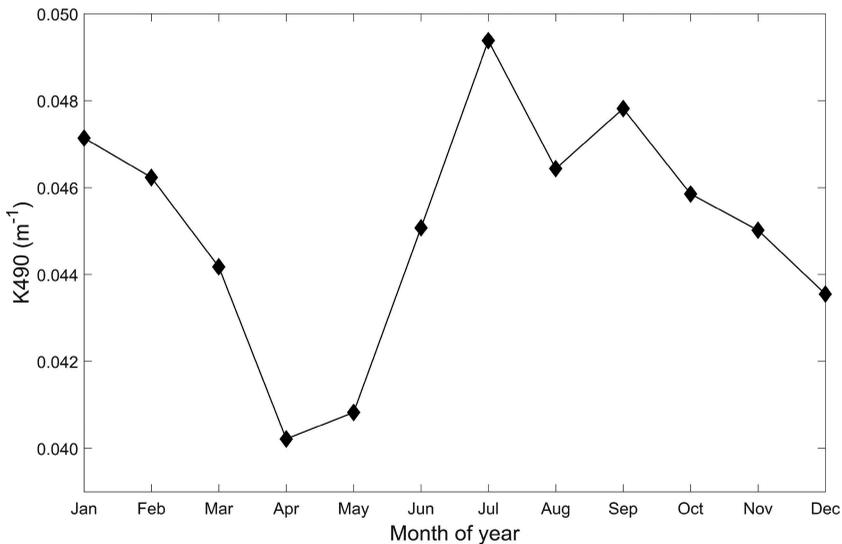


Figure 10. Sea-WiFS area-averaged time series of K490 concentration between January 2001 and December 2001.

monthly SST as measured by MODIS Aqua, and it shows a trend centred around an average of 28°C and a maximum value at approximately 29.8°C.

This secondary analysis shows that our data mining and significance test results support the hypothesis that K490 and chl-*a* are not statistically significant parameters with respect to coral disease observations. Additionally, this analysis showed a substantial influence of the concentration of UV-absorbing aerosols on both chl-*a* and K490

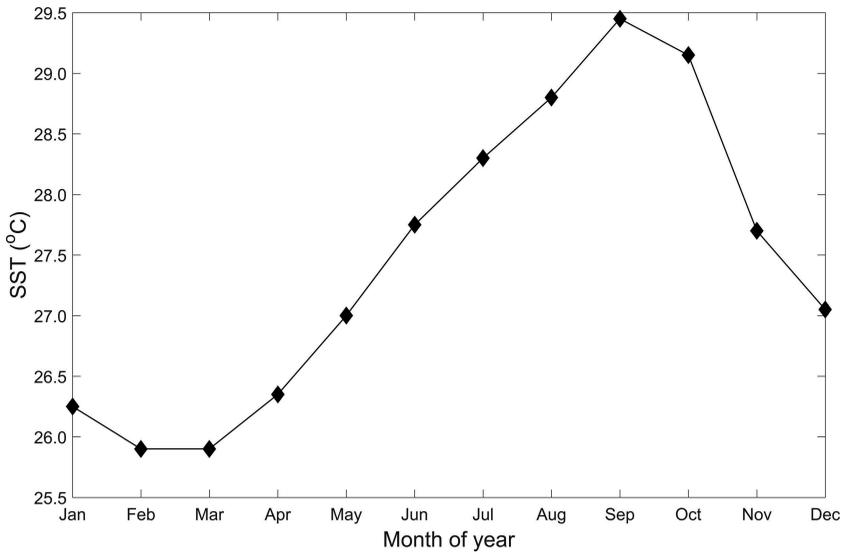


Figure 11. Area-averaged SST time series from MODIS Terra, 11 μm channel, between January 2001 and December 2001.

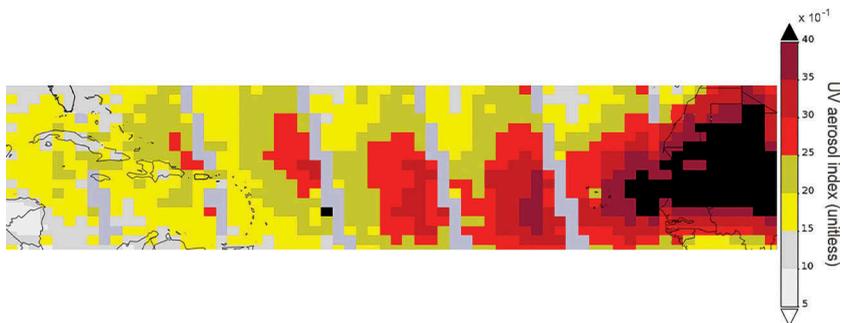


Figure 12. Map of UV AAI from TOMS EP for 1–30 June 2003.

values; however, given the data sets used for this analysis, SST and the number of coral disease observations were found only to be marginally influenced by this concentration. The difference between the area-averaged values of SST in 2001 and 2003 was approximately 1°C , and the subsequent difference between the total number of coral disease observations in these same years was approximately 48 counts, or 21%. These results, therefore, do not explicitly infer a relationship between dust concentrations and SST, but do infer a tightly coupled relationship between the number of coral disease observations and SST in the presence of dust. Therefore, given the results of the data mining algorithm, the significance test, and the secondary analysis, the influence of African dust storms on the number of coral disease observations is more dependent on ocean temperatures than on turbidity and nutrient concentration.

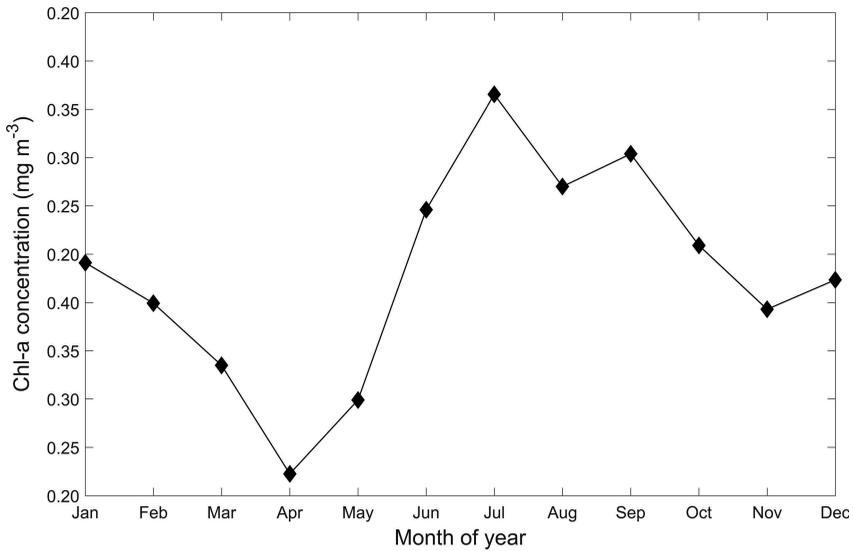


Figure 13. Sea-WiFS area-averaged time series chl-*a* concentration between January 2003 and December 2003.

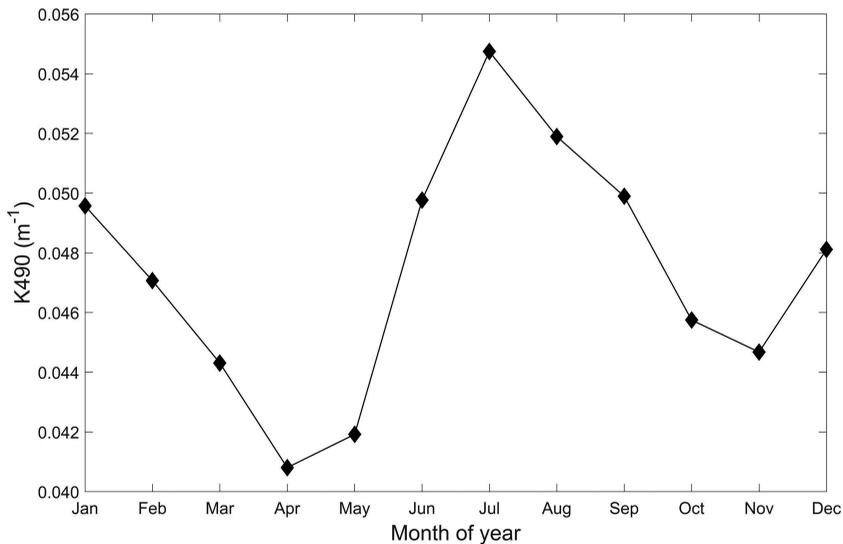


Figure 14. Sea-WiFS area-averaged time series K490 concentration between January 2003 and December 2003.

5.4. Future research

The development of models, forecasts, and new data products to monitor and track the influence of a changing climate on coral reefs remains in its infancy. Although organizations like NOAA have developed several data products to enable the comparative analyses of coral reef ecosystems, geography, environmental conditions, and

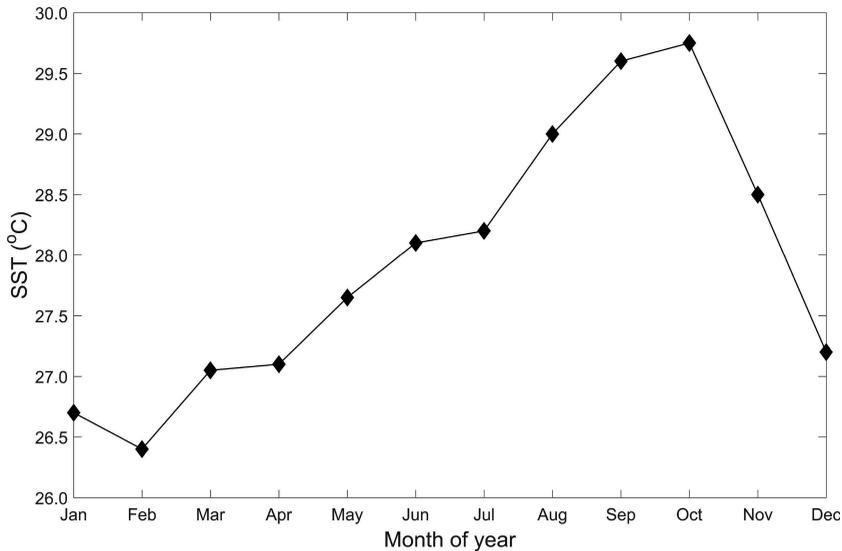


Figure 15. Area-averaged SST time series from MODIS Terra, 11 μm channel, between January 2003 and December 2003.

anthropogenic stressors for the improvement of management approaches, research and development in this field faces a few key difficulties. In this section, we present some of these difficulties with recommendations for further research.

Currently, coral disease observations are reported by divers and input manually into databases. The use of human researchers provides comprehensive and detailed classification of disease type and extent, and it is an appropriate method to retrieve accurate information. For the purpose of studying a large area such as the Caribbean Sea, however, this method suffers from low spatial and temporal resolution. As satellite data becomes ubiquitous, it is increasingly possible to overcome the resolution limitations by combining remote-sensing data with *in-situ* data; however, the process of collecting and archiving the *in-situ* data must be improved. For example, the GCDD did not have a consistent format; some entries contained detailed information while others did not. To facilitate studies of long-term environmental impacts on coral, in the Caribbean and elsewhere, research should be done to (1) impose a consistent data reporting format, (2) create a single database, such as GCDD, containing these data that are freely available to researchers, (3) develop rigorous quality assessment and control policies for this database, and (4) improve and automate the ability to access, search, and use this database.

At present, new sensors are being developed to study coral reefs. One example is the Coral Reef Airborne Laboratory (CORAL), which is a suborbital platform bearing a push-broom imaging spectrometer operating in the near UV through near infrared. This three year mission seeks to study the relationship between coral reef health and biogeophysical forcings and is a significant step in studying coral from space (NASA CORAL 2016). Other research seeks to improve classification algorithms to automatically detect diseased coral. Even with all this new available data, further research is required to develop better spatio-temporal data mining algorithms. In preparation for this study, several data

mining algorithms were considered that should be pursued in further research. Of them, the most promising is sequential pattern data mining (SDM) and SDM with time lags, and it is recommended that this study be again performed using a combination of SDM with time lags and ARDM. This is because finding instantaneous relationships between the variables presented in this study does not consider that environmental variables may influence each other over a period of time. We recommend that new data mining algorithms be developed to identify time-lagged associations among climatic variables. Together, improved remote-sensing data and the data mining algorithms used to analyse the data will contribute to improved models and forecasts of coral health.

6. Conclusion

This study supports existing hypotheses and shows not only a correlation between African dust storms and coral disease, but also that this correlation is influenced by other environmental parameters. The ARDM algorithm used in this study shows that there is a strong relationship between African dust and coral disease observations that is primarily modulated by temperature, rather than by water clarity or chlorophyll concentration. To determine the specific statistical significance of each environmental parameter with regard to its influence on the number of coral disease observations, this study uses Fisher's exact significance Test. The results of this test and a secondary, qualitative analysis support the hypothesis that coral disease in the Caribbean Sea is not influenced only by the absence or presence of African dust, but the state of the local Caribbean climate, primarily SST.

Acknowledgements

H. H. was partially supported by Exelis, Inc at the time of this research. G. C. was partially funded by the National Science Foundation (NSF) award, #1639707 and by the Office of Naval Research (ONR) award #N00014-14-1-0208 (PSU #171570).

Disclosure statement

No potential conflict of interest was reported by the authors.

ORCID

Heather Hunter  <http://orcid.org/0000-0002-1672-1000>

Guido Cervone  <http://orcid.org/0000-0002-6509-0735>

References

- Adams, A. M., J. M. Prospero, and C. Zhang. 2012. "CALIPSO-Derived Three-Dimensional Structure of Aerosol over the Atlantic Basin and Adjacent Continents." *Journal of Climate* 25 (19): 6862–6879. doi:10.1175/JCLI-D-11-00672.1.
- Andréfouët, S., P. J. Mumby, M. McField, C. Hu, and F. Muller-Karger. 2002. "Revisiting Coral Reef Connectivity." *Coral Reefs* 21 (1): 43–48. doi:10.1007/s00338-001-0199-0.
- Andréfouët, S., and B. Riegl. 2004. "Remote Sensing: A Key Tool for Interdisciplinary Assessment of Coral Reef Processes." *Coral Reefs* 23 (1): 1–4. doi:10.1007/s00338-003-0360-z.

- Baker, A. C., P. W. Glynn, and B. Riegl. 2008. "Climate Change and Coral Reef Bleaching: An Ecological Assessment of Long-Term Impacts, Recovery Trends, and Future Outlook." *Estuarine, Coastal, and Shelf Science* 80 (4): 435–471.
- Bruno, J. F., E. R. Selig, K. S. Casey, C. A. Page, B. L. Willis, C. D. Harvell, H. Sweatman, and A. M. Melendy. 2007. "Thermal Stress and Coral Cover as Drivers of Coral Disease Outbreaks." *Plos Biology* 5 (6): 1220–1227. doi:10.1371/journal.pbio.0050124.
- Burke, L., and J. Maidens. 2004. *Reefs at Risk in the Caribbean*. Washington, D.C: Word Resources Institute.
- Capone, D. G., and A. Subramaniam. 2005. "Seeing Microbes from Space." *American Society of Microbiology News* 71 (4): 179–186.
- Chauvaud, S., C. Bouchon, and R. Maniere. 1998. "Remote Sensing Techniques Adapted to High Resolution Mapping of Tropical Coastal Marine Ecosystems (Coral Reefs, Seagrass Beds, and Mangrove)." *International Journal of Remote Sensing* 19 (18): 3625–3639. doi:10.1080/014311698213858.
- Denner, E. B. M., G. W. Smith, H.-J. Busse, P. Schumann, T. Narzt, S. W. Polson, W. Lubitz, and L. L. Richardson. 2003. "Aurantimonas Coralicida Gen. Nov., Sp. Nov., the Causative Agent of White Plague Type II on Caribbean Scleractinian Corals." *International Journal of Systematic and Evolutionary Microbiology* 53: 1115–1122. doi:10.1099/ijs.0.02359-0.
- Eakin, C. M., J. A. Morgan, S. F. Heron, T. B. Smith, G. Liu, L. Alvarez-Filip, B. Baca, et al. 2010. "Caribbean Corals in Crisis: Record Thermal Stress, Bleaching, and Mortality in 2005". *Plos ONE* 5: 11. doi:10.1371/journal.pone.0013969.
- Erickson III, D. J., J. L. Hernandez, P. Ginoux, W. W. Gregg, C. McClain, and J. Christian. 2003. "Atmospheric Iron Delivery and Surface Ocean Biological Activity in the Southern Ocean and Patagonian Region." *Geophysical Research Letters* 30 (12): 1609.
- Gardner, T. A., I. M. Cote, J. A. Gill, A. Grant, and A. R. Watkinson. 2003. "Long-Term Region-Wide Declines in Caribbean Corals." *Science* 301 (5635): 958–960. doi:10.1126/science.1086050.
- Garrison, V. H., E. A. Shinn, W. T. Foreman, D. W. Griffin, C. W. Holmes, C. A. Kellogg, M. S. Majewski, L. L. Richardson, K. B. Ritchie, and G. W. Smith. 2003. "African and Asian Dust: From Desert Soils to Coral Reefs." *BioScience* 53 (5): 469–480. doi:10.1641/0006-3568(2003)053[0469:AAAFD]2.0.CO;2.
- Gasso, S., V. H. Grassian, and R. L. Miller. 2010. "Interactions between Mineral Dust, Climate, and Ocean Ecosystems." *Elements* 6 (4): 247–252. doi:10.2113/gselements.6.4.247.
- Gonzalez, N. M., F. E. Muller-Karger, S. C. Estrada, R. Perez de los Reyes, I. Victoria del Rio, and P. C. Perez. 2000. "Near-Surface Phytoplankton Distribution in the Western Intra-Americas Sea: The Influence of El Nino and Weather Events." *Journal of Geophysical Research* 105 (C6): 14029–14043. doi:10.1029/2000JC900017.
- Goudie, A. S., and N. J. Middleton. 2001. "Saharan Dust Storms: Nature and Consequences." *Earth-Science Reviews* 56 (1): 179–204. doi:10.1016/S0012-8252(01)00067-8.
- Griffin, D. W., and C. A. Kellogg. 2004. "Dust Storms and Their Impact on Ocean and Human Health: Dust in the Earth's Atmosphere." *EcoHealth* 1: 284–295. doi:10.1007/s10393-004-0120-8.
- Han, J.-W., M. Kamber, and J. Pei. 2011. *Data Mining: Concepts and Techniques*. Waltham, MA: Elsevier.
- Lausch, A., A. Schmidt, and L. Tischendorf. 2015. "Data Mining and Linked Open Data – New Perspectives for Data Analysis in Environmental Research." *Ecological Modelling* 295: 5–17. doi:10.1016/j.ecolmodel.2014.09.018.
- Lesser, M. P., J. C. Bythell, C. Gates, R. W. Johnstone, and O. Hoegh-Guldberg. 2007. "Are Infectious Diseases Really Killing Corals? Alternative Interpretations of the Experimental and Ecological Data." *Journal of Experimental Marine Biology and Ecology* 346: 36–44. doi:10.1016/j.jembe.2007.02.015.
- Mahowald, N., A. R. Baker, G. Bergametti, N. Brooks, R. A. Duce, T. D. Jickells, N. Kubilay, J. M. Prospero, and I. Tegen. 2005. "Atmospheric Global Dust Cycle and Iron Inputs to the Ocean." *Global Biogeochemical Cycles* 19 (4).
- Malone, J. 2011. *ARMADA Data Mining Tool V.-1.4*. Edinburgh, 18 February. <http://www.mathworks.com/matlabcentral/fileexchange/3016-armada-data-mining-tool-version-1-4>.

- Mennis, J., and D. Guo. 2009. "Spatial Data Mining and Geographic Knowledge Discovery—An Introduction." *Computers, Environment and Urban Systems* 33: 403–408. doi:10.1016/j.compenvurbsys.2009.11.001.
- Mennis, J., and J. W. Liu. 2005. "Mining Association Rules in Spatio-Temporal Data: An Analysis of Urban Socioeconomic and Land Cover Change." *Transactions in GIS* 9 (1): 5–17. doi:10.1111/tgis.2005.9.issue-1.
- Miller, J., E. Muller, C. Rogers, R. Waara, A. Atkinson, K. R. T. Whelan, M. Patterson, and B. Witcher. 2009. "Coral Disease following Massive Bleaching in 2005 Causes 60% Decline in Coral Cover on Reefs in the US Virgin Islands." *Coral Reefs* 28: 925–937. doi:10.1007/s00338-009-0531-7.
- Moulin, C., F. Dulac, C. E. Lambert, and U. Dayan. 1997. "Control of Atmospheric Export of Dust from North Africa by the North Atlantic Oscillation." *Nature* 387 (6634): 691–694.
- Mumby, P. J., W. Skirving, A. E. Strong, J. T. Hardy, F. L. Ellsworth, E. J. Hochberg, R. P. Stumpf, and L. T. David. 2004. "Remote Sensing of Coral Reefs and Their Physical Environment." *Marine Pollution Bulletin* 48 (3): 219–228. doi:10.1016/j.marpolbul.2003.10.031.
- NASA CORAL. 2016. *Coral Reef Research Laboratory*. Accessed November 1 2016. <https://coral.jpl.nasa.gov/>.
- NASA Goddard Space Flight Center, Ocean Ecology Laboratory, Ocean Biology Processing Group. 2014. "Reprocessing." *Sea-Viewing Wide Field-Of-View Sensor (SeaWiFS) Chlorophyll Data*. Accessed 2012. doi:10.5067/ORBVIEWS/SEAWIFS/L3B/CHL/2014.
- NOAA Earth System Research Laboratory. n.d. *NOAA Optimum Interpolation (OI) Sea Surface Temperature (SST)*. Accessed 2012. <http://www.esrl.noaa.gov/psd/data/gridded/data.noaa.oisst.v2.html>.
- Nurlidiasari, M., and S. Budhiman. 2005. "Mapping Coral Reef Habitat with and without Water Column Correction Using Quickbird Image." *Remote Sensing and Earth Sciences* 2: 45–56.
- Palandro, D., S. Andréfouët, P. Dustan, and F. E. Muller-Karger. 2003. "Change Detection in Coral Reef Communities Using Ikonos Satellite Sensor Imagery and Historic Aerial Photographs." *International Journal of Remote Sensing* 24 (4): 873–878. doi:10.1080/014311602100009895.
- Prospero, J. M., P. Ginoux, O. Torres, S. Nicholson, and T. Gill. 2002. "Environmental Characterization of Global Sources of Atmospheric Soil Dust Identified with the NIMBUS 7 Total Ozone Mapping Spectrometer (TOMS) Absorbing Aerosol Product." *Reviews of Geophysics* 40: 1. doi:10.1029/2000RG000095.
- Prospero, J. M., and R. T. Nees. 1986. "Impact of the North African Drought and El Niño on Mineral Dust in the Barbados Trade Winds." *Nature* 320: 735–738. doi:10.1038/320735a0.
- Rajasekar, U., and Q. Weng. 2009. "Application of Association Rule Mining for Exploring the Relationship between Urban Land Surface Temperature and Biophysical/Social Parameters." *Photogrammetric Engineering & Remote Sensing* 75: 385–396. doi:10.14358/PERS.75.4.385.
- Raymundo, L. J., C. S. Couch, and C. D. Harvell. 2008. *A Coral Disease Handbook: Guidelines for Assessment, Monitoring, and Management*, 120. St. Lucia: Coral Reef Targeted Research and Capacity Building for Management Program, University of Queensland.
- Reefbase. 2016. *Reefbase: A Global Information System for Coral Reefs*. Accessed 2012. <http://www.reefbase.org/main.aspx>.
- Rypien, K. L. 2008. "African Dust Is an Unlikely Source of Aspergiillus Sydowii, the Causative Agent of Sea Fan Disease." *Marine Ecology Progress Series* 367: 125–131. doi:10.3354/meps07600.
- Schutte, V. G., E. R. Selig, and J. F. Bruno. 2010. "Regional Spatio-Temporal Trends in Caribbean Coral Reef Benthic Communities." *Marine Ecology Progress Series* 402: 115–122. doi:10.3354/meps08438.
- Selig, E. R., C. D. Harvell, J. F. Bruno, B. L. Willis, C. A. Page, K. S. Casey, and H. Sweatman. 2006. *Analyzing the Relationship between Ocean Temperature Anomalies and Coral Disease Outbreaks at Broad Spatial Scales*, 111–128. *Coral Reefs and Climate Change: Science and Management* (American Geophysical Union), Washington, DC.
- Shinn, E. A., G. W. Smith, J. M. Prospero, P. Betzer, M. L. Hayes, V. Garrison, and R. T. Barber. 2000. "African Dust and the Demise of Coral Reefs." *Geophysical Research Letters* 27 (19): 3029–3032. doi:10.1029/2000GL011599.
- Steinbach, M., G. Karypis, and V. Kumar. 2000. "A Comparison of Document Clustering Techniques." *KDD 2000 Workshop on Text Mining*.

- Tan, P.-N., M. Steinbach, V. Kumar, C. Potter, S. Klooster, and A. Torregrosa. 2001. "Finding Spatio-Temporal Patterns in Earth Science Data." *KDD 2001 Workshop on Temporal Data Mining*.
- United Nations. 2016. *United Nations Environment Programme: World Conservation Monitoring Centre, Global Coral Disease Database*. Accessed 2012. <http://www.unep-wcmc.org/resources-and-data/global-coral-disease-database>.
- Webb, G. I. 2007. "Discovering Significant Patterns." *Machine Learning* 68: 1–33. doi:10.1007/s10994-007-5006-x.
- Yang, R., J. Tang, and D. Sun. 2011. "Association Rule Data Mining Applications for Atlantic Tropical Cyclone Intensity Changes." *Weather and Forecasting* 26: 337–353. doi:10.1175/WAF-D-10-05029.1.