Climate Dynamics manuscript No.
(will be inserted by the editor)

Spatial grids for hurricane climate research

James B. Elsner · Robert E. Hodges · Thomas H. Jagger

Received: date / Accepted: date

Abstract The authors demonstrate a spatial framework for studying hurricane climatology. The framework consists of a spatial tessellation of the hurricane basin using equal-area hexagons. The hexagons are efficient at covering hurricane tracks and provide a scaffolding to combine attribute data from tropical cyclones with spatial climate data. The framework's utility is demonstrated using examples from recent hurricane seasons. Seasons that have similar tracks are quantitatively assessed and grouped. Regional cyclone frequency and intensity variations are mapped. A geographically-weighted regression of cyclone intensity on SST emphasizes the importance of a warm ocean in the intensification of cyclones over regions where the heat content is greatest. The largest differences between model predictions and observations occur near the coast. The authors suggest the framework is ideally suited for comparing tropical cyclones generated from different numerical simulations.

Keywords hurricane · hexagon grid · spatial data · tessellation · geographically-weighted regression · spatial autocorrelation

J. B. Elsner
Florida State University, Tallahassee, FL USA
Tel.: 850-877-4039
E-mail: jelsner@fsu.edu

R. E. Hodges
Florida State University, Tallahassee, FL USA
Tel.: 850-877-4039
E-mail: robhodges17@gmail.com

T. H. Jagger
Florida State University, Tallahassee, FL USA
Tel.: 850-877-4039
E-mail: tjagger@blarg.net

1 Introduction

Archives of past tropical cyclones are the basis for various storm related climatologies. The frequency of cyclones at a specific location provides a temporal climatology with a focus on the occurrence rates. For example, Malmstadt et al (2010) give a climatology of severe hurricanes affecting Florida cities. The occurrence of cyclone genesis over an area provides a spatial climatology with a focus on spatial density. For example, Chand and Walsh (2009) produce a spatial climatology of cyclone origin over the Southwest Pacific Ocean.

A climatology that combines the spatial with the temporal is also useful. Elsner (2003) and Camargo et al (2007) use cluster methods on storm location at specific storm-relative events (i.e., origin and maximum intensity points) to construct track climatologies and Scheitlin et al (2010) use space-time averaging to construct a track-relative hurricane climatology. The methods are also applied to tropical cyclone records grouped by, or regressed on, climate variables (frequently for instance an index for El Niño). For example, Elsner and Jagger (2006) show that although El Niño influences the average cyclone intensity, the North Atlantic oscillation (NAO) statistically conditions the preferred cyclone track. Cyclone tracks are treated as a temporal sequence of points over space, but climate data are treated as a factor or covariate that changes over time, but without spatial dimension.

Missing is a way to combine the tropical cyclone data with the time *and* spatial dimension of the climate data. For example, although Elsner et al (2008) show that the strongest hurricanes are getting stronger as the ocean warms over the North Atlantic, is this relationship occurring over the warmest waters? This deficiency arises from a mismatch in data types; tropical cyclones are represented as points or line data, while climate is represented as gridded field data. This makes it difficult to sync the data across space and time.

Here we demonstrate an approach for combining tropical cyclone data with spatially continuous climate data. The methodology makes use of a tessellation for the cyclone basin. The corresponding grids are populated with local storm and climate information allowing further analysis and modeling. We show that hexagon grids are efficient in covering cyclone tracks and they provide a way to combine climate data with attribute data from the cyclones. We then show how local intensity relationships can be modeled using a spatial regression model. Our goal is to demonstrate the utility of a framework for combining storm and climate data. The key components of the framework are a procedure to spatially aggregate storm data and models that make use of the resulting spatial autocorrelation.

The remainder of this paper is organized as follows. Motivation for our study and the data used is outlined in Sections 2. In Section 3 we show the efficiency of hexagons tessellation over squares, and in Section 4 we analyze the frequency and intensity of activity. Seasons with similar tracks can be found based on grid overlaps. In Section 5 we show how gridded environmental data can be co-located on the grids and be used to qualitatively examine their relationship to storm intensity. In Section 6 we show how to determine

grid neighborhoods and to analyze the spatial autocorrelation. We then show how to examine the relationship between storm intensity and sea-surface temperature regionally using a geographically-weighted regression model, where the weights are based on grid neighbors. In Section 7 we provide a summary and closing remarks.

2 Motivation

2.1 Track and fields

Here we consider two different data types. Hurricane data along a track and climate data that represents a spatially continuous field. Hurricane data are obtained from the U.S. NOAA National Hurricane Center HURDAT (Jarvinen et al, 1984) and spline interpolated to one-hourly values from their native 6-hourly format (Jagger and Elsner, 2006). The use of hourly values facilitates the spatial aggregation as the average distance traveled by a hurricane in one hour is only about 18 km.

SST data comes from NOAA's Extended Reconstructed Sea Surface Temperature (version 3b) at 2° resolution, provided by the NOAA[1] (Smith et al, 2008). Upper air data comes from the 20th Century Reanalysis V2, also provided by the NOAA (Compo et al, 2006; Whitaker et al, 2004).

The standard visual presentation of a past tropical cyclone is a map displaying the track. The track consists of a connected series of locations with an intensity attribute typically attached to the location. The National Hurricane Center displays tropical cyclones in this way as part of their regular seasonal summaries.

The standard visual presentation of monthly or seasonal averages is a contour map. The averages are computed on a regular grid (e.g., latitude, longitude) and the values are contoured using interpolation. Figure 1 shows the sea-surface temperature field for August 2005 with tracks from the corresponding hurricane season overlaid. Tracks are shown only for tropical cyclones that reached hurricane intensity (33 m s^{-1}) but the track includes all cyclone intensities.

2.2 Tessellation

The map overlay is useful for summaries and exploratory analyzes. For instance, the map of the 2005 hurricane season (Fig. 1 shows a group of hurricanes over the warm Caribbean Sea and Gulf of Mexico and another, somewhat separate, group of hurricanes over the cooler central North Atlantic. However a common framework that can be used for track and field data would facilitate statistical inference and make it easier to develop statistical models.

[1] NOAA/OAR/ESRL PSD, Boulder, Colorado, USA (http://www.esrl.noaa.gov/psd/)

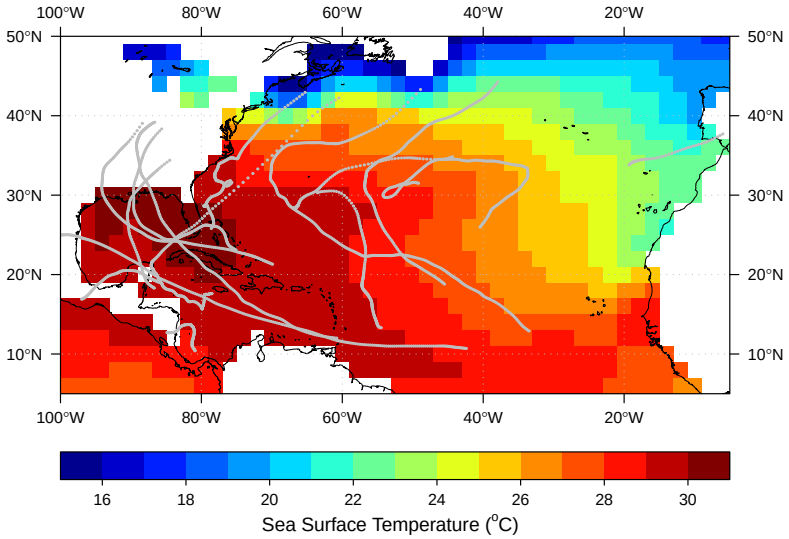

Fig. 1 Tropical cyclone tracks and sea-surface temperature. The points are 1 hr interpolations of the 6 hr observations and they indicate the paths taken by hurricanes during the 2005 season. The August 2005 temperatures are on a 2° latitude-longitude grid.

The framework we propose is a spatial lattice. Hurricane track information including location is aggregated to the same regular areas that contain climate information.

To make things simple, the lattice should cover the domain of interest completely with no overlaps or gaps, and it should do so with all grids having the same shape. In other words, the lattice should provide a tessellation of the domain. Field data on a latitude-longitude grid can be considered a tessellation. The field value at the grid point can be considered as representing an average within a rectangular area (grid) and the grids do not overlap or leave gaps. The SST map for August of 2005 (Fig. 1) is displayed as a rectangular lattice. The grids are 2° latitude by 2° longitude 'rectangles.'

While useful for visual display, this kind of rectangular lattice has limitations for further analysis. Namely, latitude-longitude grids have different areas depending on location. A 2° grid covers approximately 29 % more ocean surface centered at 10° N than it does at 40° N latitude. More important for this study, a rectangular lattice may not be the most efficient way to cover a hurricane track.

The first limitation is overcome by an area-preserving projection of the geographic coordinates onto the plane. Here a Lambert conformal (angle-preserving) conic projection with secants parallels at 23 and 38° N (centered on 60° W longitude) is used, though certainly other projections are appropriate, as well. The second limitation is addressed empirically, as discussed next.

Tessellations of the plane can be made with three simple regular geometric grids including triangles, rectangles, and hexagons. A triangle grid has three direct (straight-side) neighbors, but the neighbor grids are upside down. A rectangle grid has four direct neighbors all with the same orientation. Hexagons have six direct neighbors all with the same orientation. For a more in-depth look at tessellations see the work of Birch et al (2007), who examine rectangular and hexagonal grids for observational, experimental, and simulation studies in ecology.

For field data the grid shape is largely irrelevant, but for hurricane data it is desirable that the set of grids cover the track points. The number of grids of a given size needs to completely cover a track will depend to some extent on its shape. We define the relative efficiency as the ratio of the numbers of grids needed to completely cover a track using two different grid shapes, but each having the same area. To simplify further we disregard triangles because of the orientation issue mentioned above and focus on the efficiency of hexagonal grids relative to rectangular grids.

The use of grids to represent hurricane pathways is not new. Elsner et al (2004) examine the annual probability of hurricanes at 0.5° grid line intersections while Elsner and Kara (1999) use latitude-longitude "rectangles" to identify preferred paths for hurricanes approaching regions along the U.S. coastline similar to what was done by Hope and Neumann (1971). Brettschneider (2008) improves on this latter methodology by using hexagon grids with equal areas.

However, most studies take a different approach. Track densities and track clustering are the most common alternatives. A track density map is constructed by applying a function that smooths the set of track points over space (Xie et al, 2005; Joyner and Rohli, 2010). The smoothing kernel and kernel bandwidth need to be specified by the user and the method makes it hard to distinguish a region affected by a single, slow moving hurricane from a region affected by many quick moving hurricanes.

Track clustering is a popular alternative. Track membership in a group is determined by spatial coordinate values at specified track-relative locations. For example Elsner (2003); Elsner and Liu (2003) use the geographic coordinates of a tropical cyclone at its maximum and final hurricane (or typhoon) intensities and a k-means algorithm for determining group membership. Other clustering methods have been proposed that characterize the shape and location of tracks (Camargo et al, 2007; Nakamura et al, 2009). Usually the number of clusters needs to be set beforehand, with the procedure that can be used to estimate a "mean" track that reduces the spatial dimensions from the set of all tracks.

Dimensional reduction can help find patterns in the data, but our goal here is different. We want to extend the spatial dimension of each track so that we can geographically match the space occupied by the hurricane with the space occupied by climate information. Neither track clustering nor track density methods provide the framework we seek for combining track and field data, so we suggest a different approach.

3 Relative Efficiency of Hexagons

Having argued that grids provide a scaffolding for combining data of different types, our first step is to choose a grid shape. As mentioned, our choice is between squares and hexagons as these tessellate a plane.

Regardless of grid shape it takes fewer large grids to cover a track. Figure 2 shows the track of Hurricane Katrina (2005) using hourly points with equal-area square and hexagonal grids covering the track at two different grid resolutions. The area of the grids in the top panels are 32,807 km^2 for the squares and 32,769 km^2 for the hexagons, the area of the grids in the bottom panels are 3,276 km^2 for the squares and 3,275 km^2 for the hexagons. The small differences in grid areas arise because grids are first specified as a regular set of center grid points rather than polygons, and the initial grid center is a random value. The color corresponds to the highest hurricane intensity value of the track points inside the grid. The grids are plotted on a map using the Lambert conformal conic projection with standard parallels of 23° and 38°.

With the largest grids, the number of rectangles needed to cover Hurricane Katrina is the same as the number of hexagons. However, with the smallest grids the ratio of the number of square grids to the number of hexagon grids is 22/16 or 1.375. Thus at this scale, it takes 37.5 % more squares to cover the track of Katrina than hexagons of the same area. We say that hexagons are more efficient than squares.

Questions that arise naturally from this analysis include; are hexagons always more efficient than rectangles at covering hurricane tracks and, at what scale is the relative efficiency of hexagons the greatest? More relevant is the shortest distance through the grid between any two parallel grid sides. For a square with a unit area, the shortest distance between two parallel sides is 1 unit. For a hexagon with the same unit area, the shortest distance between two parallel sides is 1.074 units. Thus a straight-line hurricane track that intersects a grid perpendicular to a side will have a longer length through a hexagon than through a square of equal area. Thus it will take fewer hexagons stacked side-to-side to cover a straight track than squares with the same area and stacked similarly.

For curved tracks the question of relative efficiency is addressed empirically. We generate a series of tessellations with grid sizes ranging between 640 and 330,000 km^2. For each grid size we tessellate the region bounded by 8° and 55° N latitudes and 0° and 110° W longitudes using hexagons and squares.

Tessellations are constructed by first generating a set of grid center locations starting in the southwest corner of the region and then overlaying the corresponding non-overlapping grids such that each location is in the center of a grid. By randomly varying the southwest most grid center location by a small amount, and thus the rest of the locations, the number of grids that cover a hurricane track is a random variable.

Figure 3 shows the results of this procedure using grids to cover Hurricane Katrina. The grid area is plotted on the horizontal axis with a log scale and the count of grids needed to completely cover the track is given on the vertical

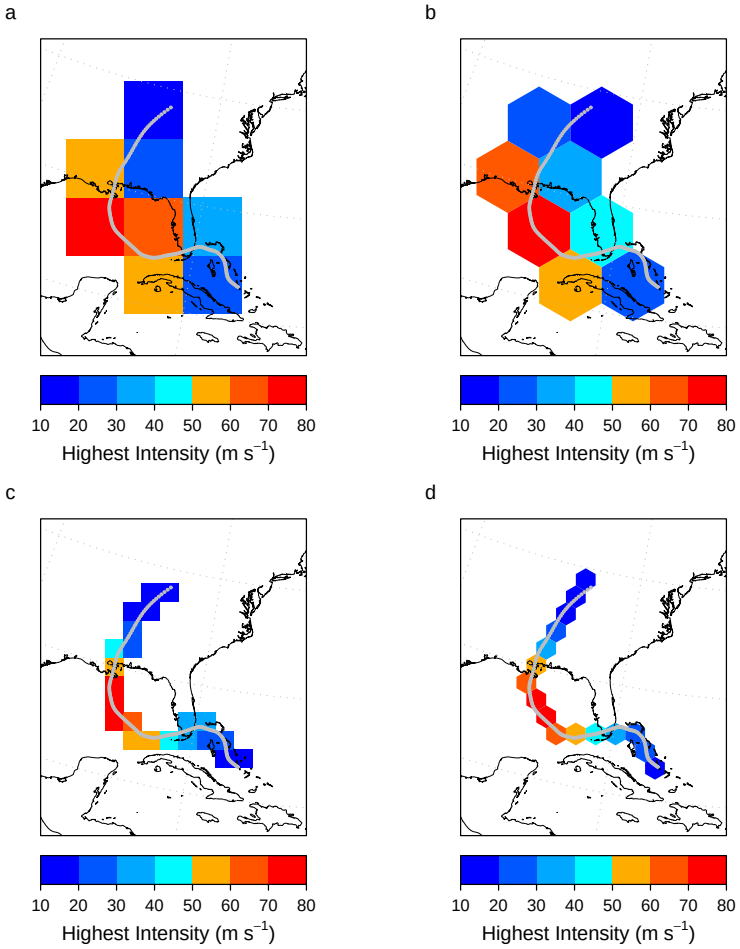

Fig. 2 Grids covering hourly track points of Hurricane Katrina (2005). The left two panels use square grids and the right two use hexagon grids. The area of the grids in the top panels are 32,807 km^2 for the squares and 32,769 km^2 for the hexagons, the area of the grids in the bottom panels are 3,276 km^2 for the squares and 3,275 km^2 for the hexagons. The color corresponds to the highest hurricane intensity value of the track points inside the grid.

axis. The individual points indicate the variation in counts as the offset is varied.

As expected, the number of grids needed to cover Katrina's track decreases with increasing grid size, and is the case for hexagons and squares. The variability in counts also decreases with increasing area. The variability arises from randomly offsetting the grid origin. The offset is chosen using a two-dimensional uniform distribution across an area equal to the grid area. For smaller grids the variability is larger for hexagons. A local regression model of the counts on grid area is shown by the curves. The fraction of the number

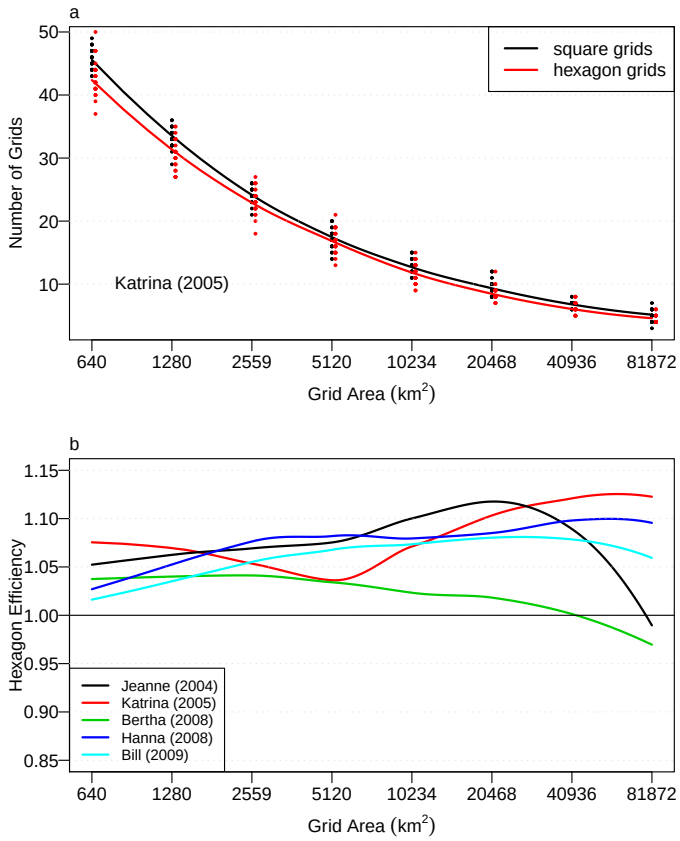

Fig. 3 (a) The number of grids needed to cover the Hurricane Katrina's track as a function of grid area using hexagon and square grids. Grid area (km^2) is given along the horizontal axis using a logarithmic scale. The dots represent the counts using random grid origin offsets and the curves are based on a local regression smoother through the points. The points representing the hexagon counts are shifted to the right slightly for clarity. (b) The ratio of the number of hexagon grids to the number of square grids as a function of grid area computed from the local regression models for Katrina and four other hurricanes. A ratio greater than unity indicates hexagons are more efficient than squares at covering the track.

of square grids to the number of hexagon grids predicted from the regressions as a function of grid area provides a metric of hexagon efficiency. A ratio of one indicates equal number of grids with values greater than one indicating it takes fewer hexagons to cover the track than squares of equal area.

The hexagon efficiency for Katrina and four other hurricanes (Jeanne 2004, Bertha 2008, Hanna 2008, and Bill 2009) are displayed. The choice of tropical storms is an attempt to get a range of track types (nearly straight to highly curved) from the set of storms over the past few years. The plot shows that hexagons tend to be more efficient than squares. The variability in efficiency is greater for larger grids. From an inspection of the plot, hexagon efficiency

tends to be high and the variability lowest at a grid area of 2559 km^2. There appears to be a limit to the efficiency near 5 % as the grid area gets smaller.

Having demonstrated the efficiency of hexagon grids relative to the standard square grids for describing hurricane tracks, next we show how hexagons can be used in the analysis of hurricane climate. We begin by examining analog seasons. We then show how climate data can be matched at the scale of the hexagons leading to new insights.

4 Seasonal Activity

4.1 Analog seasons

It is often of interest to consider which hurricane seasons are similar. Similarity can be measured by the number of hurricanes or by the number of landfalls or a combination of both. For instance, 1986 and 2003 are similar as they each featured one hurricane hitting the U.S. Gulf coast and one hurricane hitting the U.S. East coast. There were also no Florida hurricanes and no major hurricane land falls in these years.

A generalization of this kind of analog is achieved by matching seasonal hurricane tracks. Given a season of hurricane tracks, what other seasons had a similar set of tracks? With tracks themselves, the problem is rather complex. However, by converting the track to grids, the problem is simplified. Figure 4 shows four seasons using hexagonal grids. A single grid has an area of 2559 km^2.

A grid is red if contains at least one track point and white otherwise. For a given set of red and white grids defining a season of hurricane tracks, we can use a metric that quantifies the set's similarity with a set of grids from another season. A simple metric is Jaccard's similarity coefficient, which is defined as the size of the intersection divided by the size of the union of the sets.

Given two seasons each with n grids, the Jaccard coefficient is a measure of the overlap in occurrences. A grid is labeled either 1 or 0 depending on whether it contains at least one track point. Let M_{11} be the number of grids that have 1's in two different seasons, M_{01} be the number of grids that have a 0 in the first season and 1 in the second season, and M_{10} be the number of grids that have a 1 in the first season and 0 in the second season. The Jaccard's coefficient, J, is given by

$$J = \frac{M_{11}}{M_{01} + M_{10} + M_{11}} \tag{1}$$

Values of J range between 1 when all grids match to 0 when no grids match and is undefined for two years without hurricanes. We compute J between all pairs of seasons from 1943 through 2009 (Fig. 5). At the grid scale used, the average value of J is 0.1 indicating very little overlap. The largest match ($J = 0.28$) occurs between the 1946 and 1968 hurricane seasons. The 1946 season also matches the 1970 season and the 1955 season matches the 1999

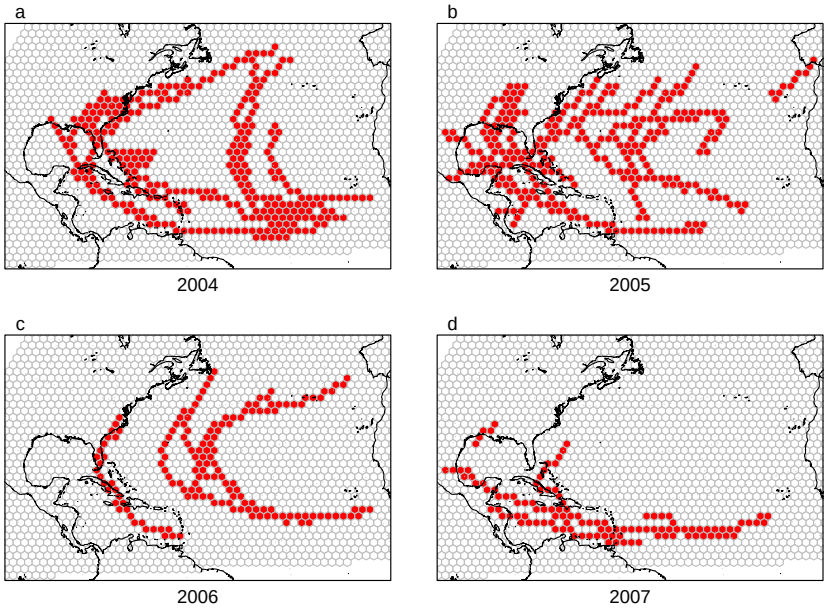

Fig. 4 Seasonal track grids. A grid is red if it contains at least one track point and white otherwise. Track points are hourly locations of the tropical cyclone at all intensities for a cyclone that reached hurricane intensity during its lifetime.

season ($J = 0.27$). Seasons that match can be considered analogs in terms of local occurrence.

The analysis can continue by grouping years based on their pairwise J values or creating a network of links with the strength of the link provided by J. The topology of the network, which might provide new insight into hurricane climatology (e.g., Elsner et al (2009)), can than be examined.

In the bottom panel the order of years is based on clustering the Jaccard coefficient (actually 1 minus J, since J is not a proper distance metric). Years with similar J values are grouped using a hierarchical clustering routine. Initially each year is assigned to its own cluster and then the algorithm proceeds to join the two most similar clusters, continuing until there is a single cluster. At each merge the distances between clusters is recomputed based on minimizing the within-cluster distances and maximizing the between-cluster distances.

Similarity in years, as measured by J and computed from the gridded hurricane paths, is largely a function of basin-wide frequency. Years with many

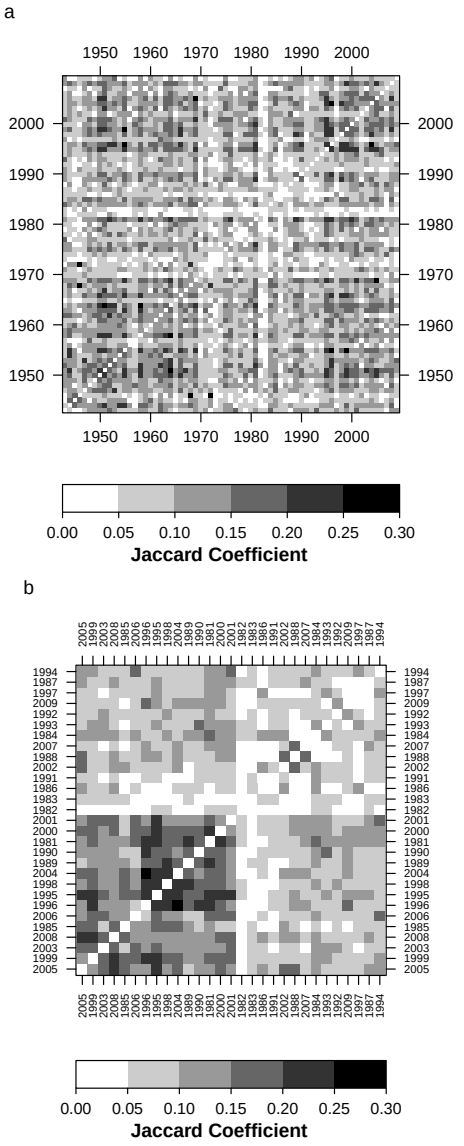

Fig. 5 Jaccard coefficients based on hurricane paths. The coefficients are computed based on the set of grids defining hurricane paths for each season. Higher values indicate years with greater similarity in paths. The lower panel shows the values only for years since 1981 and the years are grouped by a clustering of the Jaccard coefficients.

hurricanes tend to have similar values of J, although there are exceptions. For example, 2006 with only 5 hurricanes is grouped with the other, more active, years. This is because like the more active year, hurricanes in 2006 occurred over the central Atlantic near Bermuda.

4.2 Frequency and intensity

Along with locational information, the grids can be populated with other hurricane information. For instance, Fig. 6 shows the geographic distribution of tropical storm and hurricane frequency and maximum intensities for the 2005 hurricane season using hexagons (Elsner and Jagger, 2010). Here we set the grid size to 37,882 km^2, which generates a set of 173 hexagons across the domain. The larger grid size is used to increase the number of track points per grid. Only those 54 grids that contain at least one track point are used as can be seen by the hourly storm points.

The highest concentration of activity, as measured by the number of storm hours, occurs over the western Caribbean Sea into the Gulf of Mexico, and over the central North Atlantic, near Bermuda. The yellow hexagon grid over Bermuda contains exactly 200 track points. The bottom panel shows the maximum intensity using the track points within each grid. The highest intensities are occur over the western Caribbean into the Gulf of Mexico and offshore of the U.S. East coast.

The correlation between local activity level and maximum intensity is 0.61 indicating some relationship between environmental factors that control tropical cyclone frequency and factors that control intensity (Elsner and Jagger, 2010). Although the relationship is not tight as the maximum storm intensity in the grid with the most activity (over Bermuda) is only 40 m s^{-1}.

5 Environmental Variables

An advantage to representing tracks with grids is that climate information given in the form of field data can be attached to the same grids. Here we show an example using sea-surface temperature (SST) and upper-air temperature (UAT) from July 2005. Most hurricanes occur between August and October, so July values indicate conditions immediately prior to start of the most active part of the season.

As mentioned in §2 climate data are typically provided at the intersections of regular latitude and longitude grids. These data can be averaged locally within the grids represented by the season's hurricane tracks. Grids without hurricanes are left blank.

At 25° N latitude the 2° latitude/longitude SST grids cover an area of approximately the same size as the hexagon grids used in the previous section. The SST locations are converted to planar coordinates using the same Lambert conformal conic projection. Here we aggregated the track and environmental

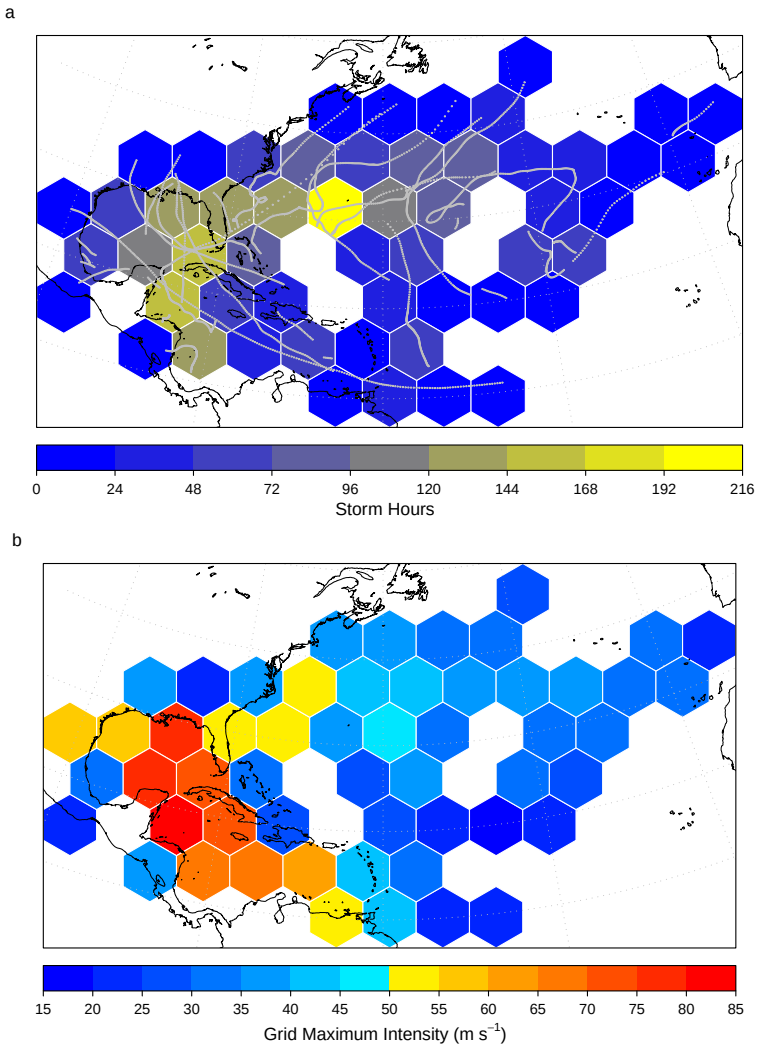

Fig. 6 Local frequency and intensity of tropical storms and hurricanes during the 2005 season. The number of storm hours is given in the top panel and the maximum intensity over all hours is given in the bottom panel.

data onto a common hexagon grid that tessellates the domain. Grids that do not contain track points are removed. Environmental values at native grid locations are averaged within the hexagon grid.

Figure 7 shows the SST and UAT maps for the 2005 hurricane season. As expected grids with the highest SST values are over the Gulf of Mexico extending into the Caribbean Sea. Average temperatures in these grids exceed 28° C. There are a few grids over the northern part of the basin where storms are collocated with lower SST values. But these grids have relatively warm

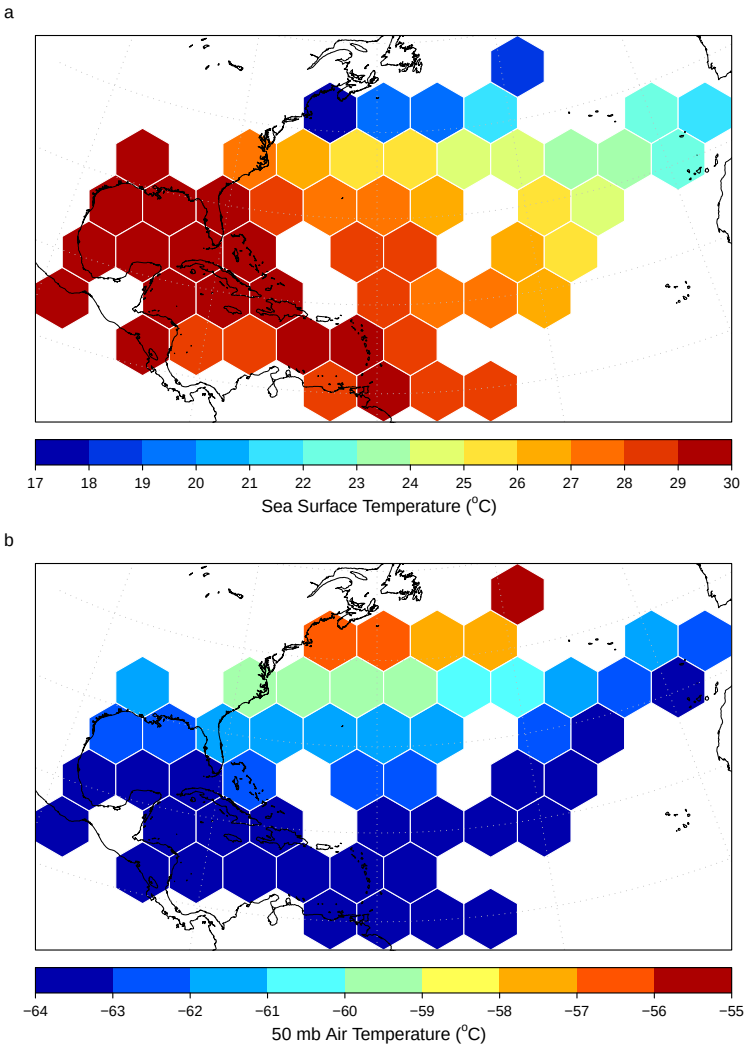

Fig. 7 SST and 50 hPa temperature during July 2005. (a) SST values (°C) averaged in the hexagon grids for July 2005. (b) Upper air temperature (50 hPa) temperature averaged in the same hexagon grids for July 2005.

UAT, with the coldest upper-air grids over the southern Gulf of Mexico into the Caribbean Sea.

Once the data are prepared in this manner, the climatologist has various options for analyzing and modeling them. For example, Fig. 8 shows the same grids colored by groups where the groups are defined by a two-way table of storm intensity and SST. We define groups by above and below median values of the variables. The median SST and storm intensity values are 28° C and 36.9 m s^{-1}, respectively.

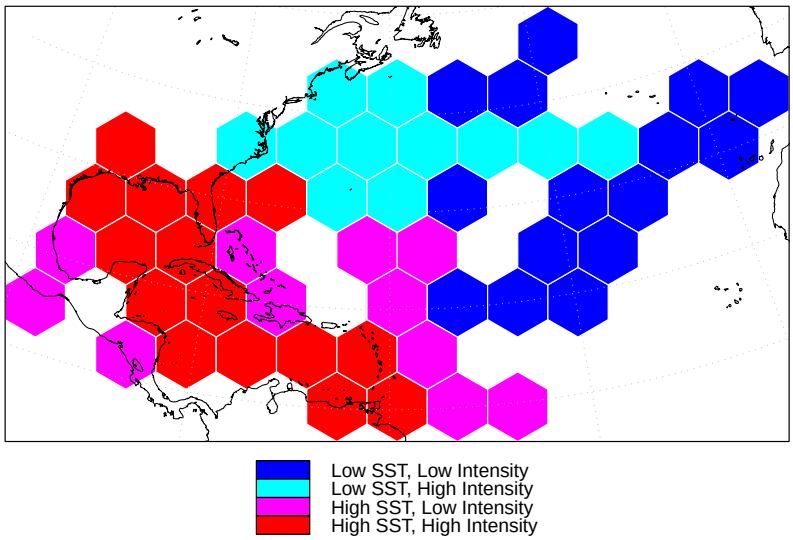

Fig. 8 SST and tropical cyclone intensities for 2005. Groups of SST and tropical cyclone intensity based on basin-wide median values. Blue indicates grids with below median values of SST and storm intensities, cyan indicates grids with below median values of SST and above normal storm intensities, magenta indicates grids with above median values of SST and below median values of storm intensities, and red indicates grids with above median values of SST and storm intensities.

The red colored grids indicate regions of high intensity and relatively high ocean temperature and the blue colored grids indicate regions of low intensity and relatively low ocean temperature. More interesting are the regions where the variables do not match. In the magenta colored grids low storm intensity is coupled with relatively high ocean temperature indicating storms are below their thermodynamic potential. In contrast, in the cyan colored grids high storm intensity is coupled with relatively low temperature indicating storms above their potential.

As the above shows, the spatial grids provide a framework that can be used to analyze relationships between hurricane characteristics and climate regionally. A more quantitative approach is to use statistical methods. Next we quantify the spatial correlation in the grid of storm intensities and use a spatial regression model to quantify the change in intensity as a function SST locally.

6 Spatial Correlation and Spatial Regression

The real value in using a common data framework comes with the ease at which relationship can be quantified locally. Local relationships are sharpened by the spatial correlation among values across the grids. Here we show how to

quantify this spatial correlation and make inferences about it and then show an example of how to quantify the relationship between storm intensity and SST locally.

6.1 Spatial autocorrelation

Values obtained by aggregating observations within contiguous grids will, in general, be self (auto) correlated. That is; values in neighboring grids will tend to be more similar than values in grids farther away. Like temporal autocorrelation, spatial autocorrelation is about this proximity, but it is more complex because of the extra dimension. One measure of spatial autocorrelation for grid data is called Moran's I (Moran, 1950). Although not used widely in weather and climate studies, de Beurs and Henebry (2008) use Moran's I to identify eco-regions and biomes with significant spatial autocorrelation within vegetative land surfaces based on an index of the North Atlantic Oscillation.

Moran's I is defined as

$$I = \frac{N}{s_o}\frac{y^T W y}{y^T y} \tag{2}$$

where N is the number of grids, y is the vector of values on the grids where the values are deviations from the mean, W is a weights matrix and the subscript T indicates the vector transpose operator.

The spatial weights matrix W with the number of rows equal to the number of columns equal to the number of grids is constructed by first identifying neighboring grids and then assigning them weights. Neighbors are identified using distances or contiguity. Figure 9 shows the network of neighbors for our hexagon tiling using contiguity. With contiguity, grids that share a common border are neighbors. For our hexagon tiling the contiguity neighbors are the same as the nearest distance neighbors.

The average number of neighbors across the tiling (set of grids) is 4.07. Thirty-seven percent of the neighbors are connected to 4 other neighbors. With the neighborhood defined using contiguity, we assign equal weights to each of the neighbors. For example, the grids with six neighbors (7 of them) each neighbor gets a weight of 1/6.

Since the y values in the above formula are deviations from the regional mean, I is equivalent to a regression coefficient in a regression of Wy on y. Thus we can visualize the linear association between the values of y at a grid and the values of y averaged over the grid's neighborhood using a scatterplot and a least-squares regression line through the points as in Fig. 10. The Moran's scatterplot for the set of maximum wind intensities suggest a fairly large amount of spatial autocorrelation with grids having large values of intensity surrounded by grids also with large values of intensity and vice versa. The slope coefficient is 0.54.

A test of the significance of this value against the null hypothesis of a zero slope (no spatial autocorrelation) is made by randomizing the intensity values across the domain and recomputing the value of I. That is, we preserve the

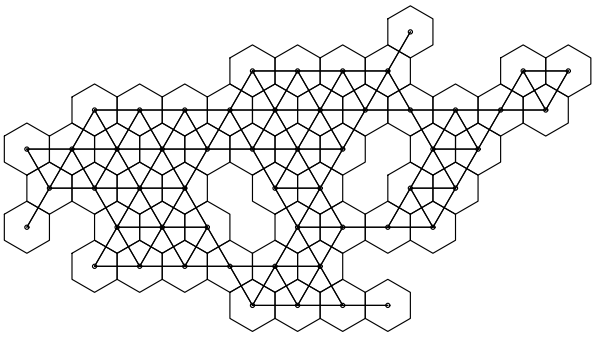

Fig. 9 Neighborhood network for the hexagon tiling. The grids cover the hurricane and tropical storm tracks from 2005. The grid centers are connected only if a grid shares a border with another grid (spatial contiguity).

neighborhood definitions given in W, but we randomly assigned an intensity value to each grid where the assignment is based on choosing an intensity value from the set of 52 values. In this way there is no spatial autocorrelation and the distribution of the set of I values computed in this way provides a bootstrapped estimate of the p value as evidence in support of the null hypothesis. The results of this procedure applied 1000 times is shown in Fig. 10 along with the value of Moran's I computed using the actual configuration of intensity values. It is clear from the results that there is non-zero spatial autocorrelation with a significance level of near 0.001.

6.2 Geographically-weighted regression

Spatial models incorporate spatial autocorrelation. This makes the model parameters stable and statistical tests more reliable. For instance, confidence intervals on a regression slope from a spatial regression model will have the proper coverage probabilities and the prediction errors will be smaller compared with the non-spatial alternative.

Spatial dependency can enter the regression model directly by adding a lagged variable to the model or by including a spatially correlated error term (Anselin et al, 2006). Spatial dependency can also enter the spatial model by allowing the relationship between the response and the explanatory variable to vary across the tiling. This is called geographically-weighted regression (Brunsdon et al, 1998; Fotheringham et al, 2000) and the model parameter values will vary depending on the grid location.

This is accomplished by weighting the grid values near the grid of interest more than the grid values farther away. For example, a regression of storm

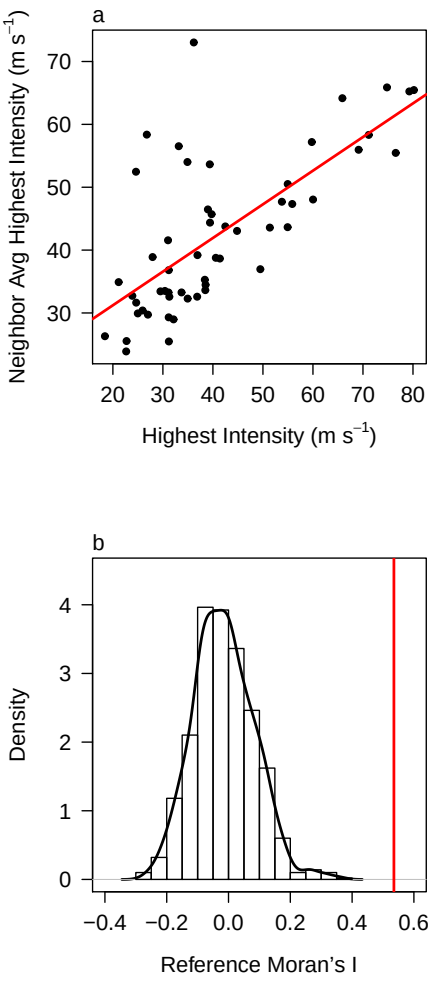

Fig. 10 Moran's spatial autocorrelation. (a) Scatterplot of the intensity (highest over all observations inside the grid) versus the neighborhood average highest intensity. The neighborhoods are defined by spatial contiguity. The least-squares linear slope is the value of Moran's I as a measure of spatial autocorrelation. (b) Results of a Monte Carlo simulation showing the significance of the spatial autocorrelation in storm intensities.

intensity on SST is performed using the paired intensity and SST values at each of the 54 grids. For each grid, the weight associated with a paired value in another grid is inversely proportional to distance between the grids. In this way the relationship between intensity and SST is localized. Geographically-weighted regression where the localization occurs in geographic space is equiv-

alent to a local linear regression where the localization occurs in the space of explanatory variables.

The classic regression model consists of a vector y of response values and a matrix X containing the set of explanatory variables plus a row vector of one's. The relationship is modeled as

$$y = X\beta + \varepsilon \tag{3}$$

where β is a vector of regression coefficients and $\varepsilon \sim N(0, \sigma^2 I)$ is a vector of independent and identically distributed residuals with variance σ^2. The maximum likelihood estimate of β is given by

$$\hat{\beta} = (X^T X)^{-1} X^T y. \tag{4}$$

With geographically-weighted regression the relationship between the response vector and the explanatory variables is

$$y = X\beta(g) + \varepsilon \tag{5}$$

where g is a vector of geographic locations, here the set of hexagons with storm observations and

$$\hat{\beta}(g) = (X^T W X)^{-1} X^T W y \tag{6}$$

where W is, again a weights matrix, but this time it is given by

$$W = \exp(-d^2/2h^2) \tag{7}$$

where d is a matrix of pairwise distances between the grids and h is the bandwidth. Here the bandwidth is a constant value and is chosen based on a cross-validation procedure (Brunsdon et al, 1998).

Our interest is storm intensity as a function of SST, but the highest storm intensity within the grid will also depend on the number of observations. In general a grid with a larger number of storm hours will have a higher intensity. Thus our spatial model includes SST and storm hours as explanatory variables in which case the SST coefficient from the regression is the marginal effect on intensity after accounting for storm hours.

Figure 11 shows the results. The grids are colored according to the value of the SST coefficient. The SST coefficient represents a local "trend" of intensity as a function of SST holding storm count constant. Hexagons with positive coefficients indicating a direct relationship between storm strength and ocean warmth in m $s^{-1}/°C$ are displayed in yellow to red colors and those with negative coefficients are shown with blue hues. Grids with the largest positive coefficients (greater than 4 m $s^{-1}/°C$) are found over the Caribbean Sea extending into the southeast Gulf of Mexico and east of the Lesser Antilles. Positive coefficients extend over much of the Gulf of Mexico and northeastward up the eastern seaboard. A region of negative coefficients (greater than -4 m $s^{-1}/°C$) is noted over the central North Atlantic.

Local significance of the coefficients can be assessed by dividing the SST coefficient by its standard error. The ratio, called the t-value, has a t distribution under the null hypothesis of a zero coefficient value. Regions of high

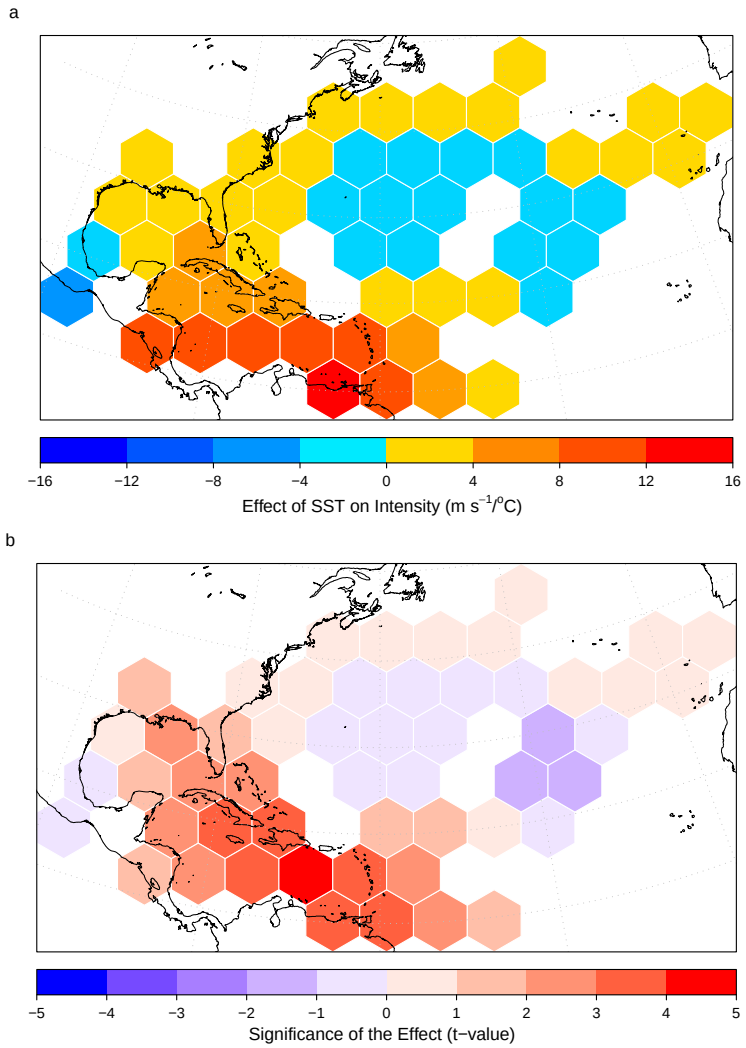

Fig. 11 Marginal effect of SST on storm intensity (a) SST coefficient of a geographically-weighted regression of storm intensity on SST and storm hours. The bandwidth for the regression is 256.5 km. (b) The ratio of the trend to the standard error of the trend as a measure of statistical significance. Values greater than an absolute value of 2 are considered significant above a trend of zero and are shaded using an inverted color transparency level. Data are for the 2005 hurricane season.

t-values (absolute value greater than 2) denote statistical significance and generally correspond with regions of high upward trends including most of the Caribbean sea and the eastern Gulf of Mexico.

Arbitrary grid boundaries are a problem when small changes lead to different results (modifiable areal unit problem) so a good strategy is to rerun the model using grids with larger or smaller areas. Here we decrease the grid

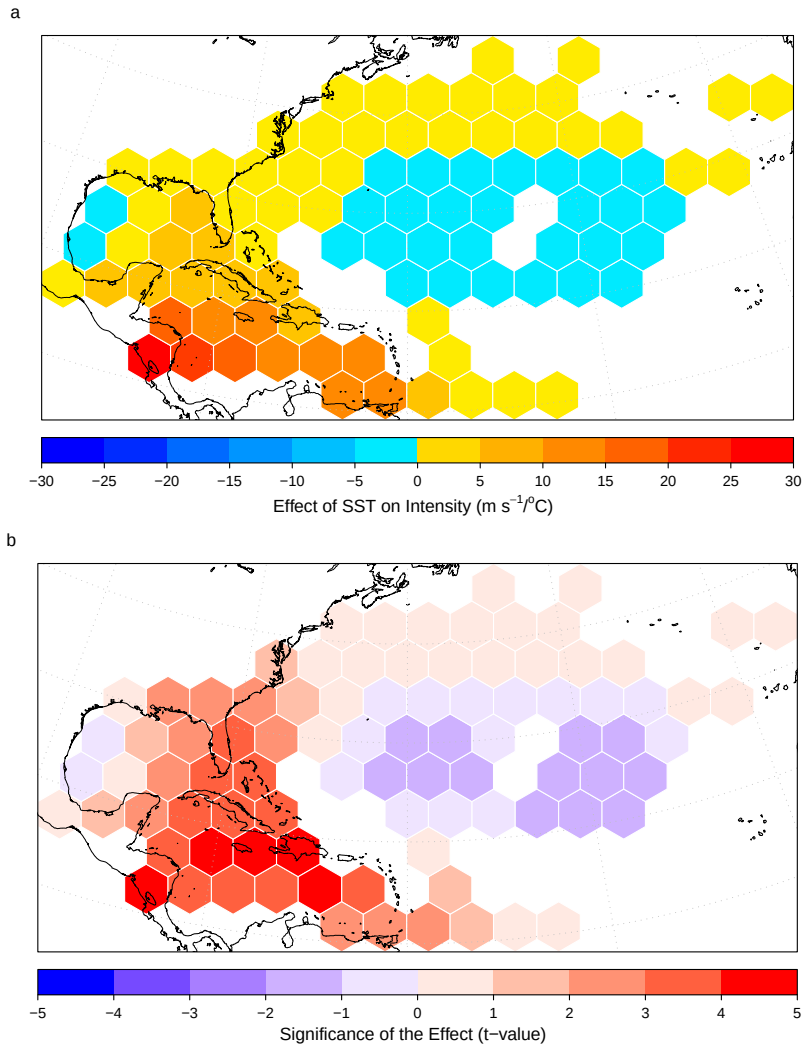

Fig. 12 Marginal effect of SST on storm intensity. Same as Fig. 11, but with smaller grids. (a) SST coefficient of a geographically-weighted regression of storm intensity on SST and storm hours. (b) The ratio of the trend to the standard error of the trend as a measure of statistical significance.

area by 50 %. This increases the number of grids occupied by storms and SST increases from 52 to 89. We then rerun the regression model and display the result in Fig. 12. The pattern is similar, featuring large and statistically significant positive coefficients in grids over the Caribbean and Gulf of Mexico.

To assess model adequacy, we examine the residuals. A residual is the difference between the intensity observed in the grid and intensity predicted by the geographically-weighted regression model. Positive residuals indicate

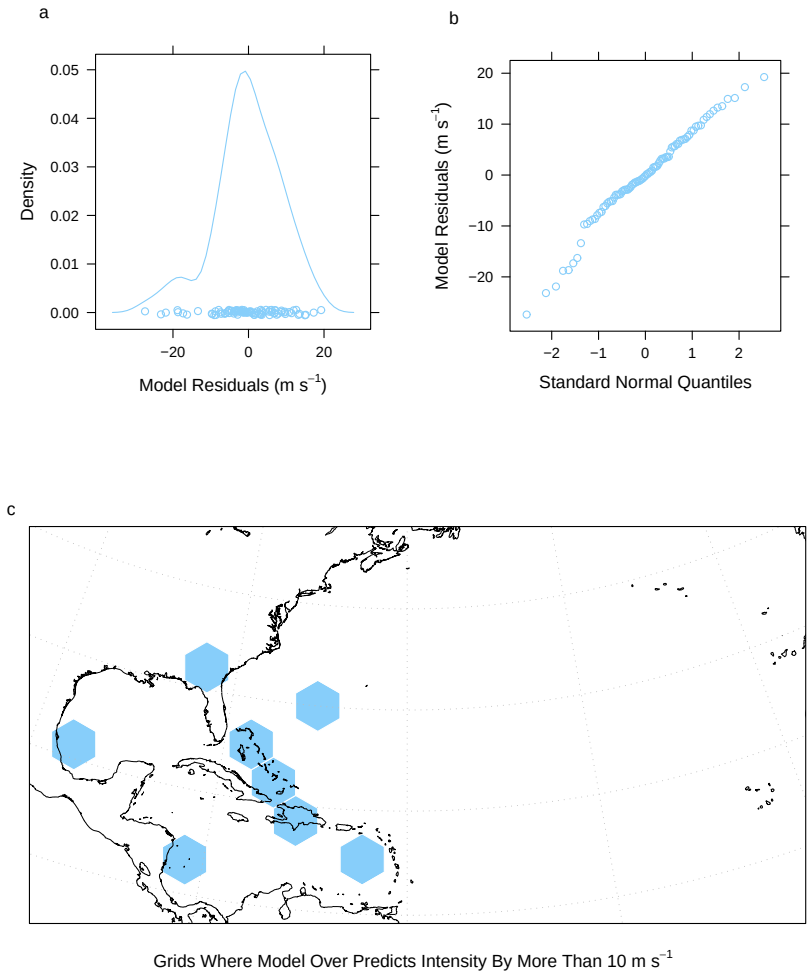

Fig. 13 Model residuals. (a) Distribution of the residuals using a smoothed kernel density where the bandwidth is 75 % of the data. (b) Scatter plot of the residuals against a set of standard normal variates. (c) Location of the eight largest negative residuals (>10 m s^{-1}). Negative residuals indicate the model over predicts storm intensity.

the model under predicts the intensity given the storm hours and SST and negative residuals indicate the model over predicts intensity. Figure 13 displays the residuals.

The residuals are centered on zero with a symmetric distribution, although there are about eight negative outliers. These outliers can be seen on the quantile-normal plot as they fall below the straight line formed by the majority of the residuals. Thus the residuals are approximately normally distributed,

but the left tail of the distribution is somewhat long. The corresponding map shows the location of the grids where the model over predicts regional intensity by more than 10 m s^{-1}. The residual map reveals that large over prediction in most cases occurs in grids near or over land. This makes sense; in these regions although SST values are high, the storm intensities are limited to some extent by influences associated with land interaction processes. The model might be improved by adding a factor indicating the presence or absence of land inside the grid or a covariate indicating the proportion of grid that is covered by land.

Other spatial regression models can be employed with this framework. For instance if interest is on cyclone counts or the presence/absence of cyclones then a Poisson or logistic model can be specified (Jagger et al, 2002). Moreover, if interest is primarily prediction rather than explanation, a spatial regression model that includes a term for the spatial autocorrelation can be specified either through a correlated error term or through a lagged response variable.

7 Summary

Tropical cyclone climatology will advance with models that combine the intrinsic space-time nature of cyclone tracks with climatic field data. This paper illustrates the use of hexagon grids to accomplish this merger.

We begin by comparing equal-area square and hexagon tessellations of hurricane tracks. As expected from differences in their geometry, we find hexagons about 7 % more efficient at covering the track paths compared with squares. We then show how hurricane seasons can be represented by such grids. A metric defining the amount of grid overlap between two seasons is used to search for seasonal track analogs and a clustering procedure employed to group years based on this metric. Of all pairwise sets, the 1946 and 1968 seasons have the most track overlap. The method could be used to identify regional shifts in storm occurrences that might help shed light on observational biases (Landsea et al, 2010; Villarini et al, in press).

We then show storm attribute data on the grids including the total storm hours and maximum intensity. The analysis provides a raster-like snapshot of the storm season. We argue that although the grids make the snapshot less focused than a track map, the loss of information at the track scale is compensated by the ability to add co-located climate data. We show an example with SST data averaged in the grids.

Importantly we show that by creating a single scaffolding for hurricane track attributes and climate field data, spatial modeling is greatly facilitated. This is demonstrated with a geographically-weighted regression of maximum storm intensity on SST and storm hours. We show statistically significant and large upward trends in regional storm intensities conditional on SST for the 2005 season. Trends are most pronounced and largely statistically significant for the greater part of the Caribbean Sea and the Gulf of Mexico (Elsner and

Jagger, 2010). An analysis of model residuals provides evidence that storm location relative to land might improve model predictions.

We leave for a later a more comprehensive climatology. Our goal here was to show how to combine storm and climate data in a way that makes further analysis and modeling at the regional scale possible. The implications could be far reaching. Hurricane climate studies are hampered by the relatively short historical archive of storms. The framework described in this paper is a way to "extend" this archive. By analyzing and modeling regional hurricane activity we effectively increase sample size. For example, we show how a spatial regression on data from a single season can help us better understand the relationship between SST and intensity. The keys are a method to spatially aggregate storm data and models that make explicit use of the resulting spatial autocorrelation. The methods could also be used in conjunction with other techniques for utilizing historical hurricane chronologies (Scheitlin et al, 2010).

Finally we note that the methodology provides a way to combine numerical model output with observations. Indeed, output from a general circulation model could be respecified from latitude/longitude coordinates to equal-area hexagons, even if the computations are not done on a geodesic grid (Randall et al, 2000). This would also provide a basis for comparing storm climatologies generated from different models.

All analysis and modeling was done using the open-source R package for statistical computing with much of the work done using the new S4 classes for spatial data developed by Bivand et al (2008). The code and data used to produce the figures in this paper are available on our website. We encourage readers to learn R and to use the code to expand on our initial results.

Acknowledgements The work was supported with contracts from the U.S. NSF under grant ATM0738172 and from the Strategic Environmental Research and Development Program under grant SERDP SI-1700.

References

Anselin L, Syabri I, Kho Y (2006) GeoDa: An introduction to spatial data analysis. Geographical Analysis 38(1):5–22, Symposium on Spatial Information Science for Human and Social Science, Tokyo, JAPAN, JAN, 2004

de Beurs KM, Henebry GM (2008) Northern annular mode effects on the land surface phenologies of northern Eurasia. Journal of Climate 21(17):4257–4279, DOI 10.1175/2008JCLI2074.1

Birch CP, Oom SP, Beecham JA (2007) Rectangular and hexagonal grids used for observation, experiment and simulation in ecology. Ecological Modelling 206(3-4), DOI 10.1016/j.ecolmodel.2007.03.041

Bivand R, Pebesma E, Gomez-Rubio V (2008) Applied Spatial Data Analysis with R. Springer

Brettschneider B (2008) Climatological hurricane landfall probability for the United States. Journal of Applied Meteorology and Climatology 47(2):704–716, DOI 10.1175/2007JAMC1711.1

Brunsdon C, Fotheringham S, Charlton M (1998) Geographically weighted regression - modelling spatial non-stationarity. Journal of the Royal Statistical Society Series D-The Statistician 47(Part 3):431–443, Workshop on Local Indicators of Spatial Association, Leicester, England, Jun, 1996

Camargo SJ, Robertson AW, Gaffney SJ, Smyth P, Ghil M (2007) Cluster analysis of typhoon tracks. Part I: General properties. Journal of Climate 20(14):3635–3653, DOI 10.1175/JCLI4188.1

Chand SS, Walsh KJE (2009) Tropical cyclone activity in the fiji region: Spatial patterns and relationship to large-scale circulation. Journal of Climate 22(14):3877–3893, DOI 10.1175/2009JCLI2880.1

Compo GP, Whitaker JS, Sardeshmukh PD (2006) Feasibility of a 100-year reanalysis using only surface pressure data. Bulletin of the American Meteorological Society 87(2):175+, DOI 10.1175/BAMS-87-2-175

Elsner JB (2003) Tracking hurricanes. Bulletin of the American Meteorological Society 84(3):353–356, DOI 10.1175/BAMS-84-3-353

Elsner JB, Jagger TH (2006) Prediction models for annual US hurricane counts. Journal of Climate 19(12):2935–2952

Elsner JB, Jagger TH (2010) On the increasing intensity of the strongest Atlantic hurricanes. In: Elsner JB, Hodges RE, Malmstadt JC, Scheitlin KN (eds) Hurricanes and Climate Change, volume 2, Springer, New York, chap 10, DOI 10.1007/978-90-481-9510-7

Elsner JB, Kara AB (1999) Hurricanes of the North Atlantic: Climate and Society. Oxford University Press

Elsner JB, Liu Kb (2003) Examining the ENSO-typhoon hypothesis. Climate Research 25(1):43–54

Elsner JB, Niu XF, Jagger TH (2004) Detecting shifts in hurricane rates using a Markov chain Monte Carlo approach. Journal of Climate 17(13):2652–2666

Elsner JB, Kossin JP, Jagger TH (2008) The increasing intensity of the strongest tropical cyclones. Nature 455(7209):92–95, DOI 10.1038/nature07234

Elsner JB, Jagger TH, Fogarty EA (2009) Visibility network of United States hurricanes. Geophysical Research Letters 36, DOI 10.1029/2009GL039129

Fotheringham AS, Brunsdon C, Charlton M (2000) Quantitative Geography: Perspectives on Spatial Data Analysis. Sage

Hope J, Neumann C (1971) Digitized atlantic tropical cyclone tracks. Tech. Memo to the National Weather Service, National Oceanic and Atmospheric Administration

Jagger TH, Elsner JB (2006) Climatology models for extreme hurricane winds near the United States. Journal of Climate 19(13):3220–3236

Jagger TH, Niu XF, Elsner JB (2002) A space-time model for seasonal hurricane prediction. International Journal of Climatology 22(4):451–465, DOI 10.1002/joc.755

Jarvinen BR, Neumann CJ, Davis MAS (1984) A tropical cyclone data tape for the north atlantic basin, 1886–1983: Contents, limitations, and uses. Technical Memo. 22, NOAA NWS NHC

Joyner TA, Rohli RV (2010) Kernel density estimation of tropical cyclone frequencies in the North Atlantic basin. International Journal of Geosciences 1:121–129, DOI 10.4236/ijg.2010.13016

Landsea CW, Vecchi GA, Bengtsson TR Lennart and (2010) Impact of duration thresholds on Atlantic tropical cyclone counts. Journal of Climate 23(10):2508–2519, DOI 10.1175/2009JCLI3034.1

Malmstadt JC, Elsner JB, Jagger TH (2010) Risk of strong hurricane winds to Florida cities. Journal of Applied Meteorology and Climatology 49:2121–2132

Moran PAP (1950) Notes on continuous stochastic phenomena. Biometrika 37:17–33

Nakamura J, Lall U, Kushnir Y, Camargo SJ (2009) Classifying North Atlantic tropical cyclone tracks by mass moments. Journal of Climate 22(20):5481–5494, DOI 10.1175/2009JCLI2828.1

Randall D, Heikes R, Ringler TD (2000) General Circulation Model Development: Past, Present, & Future, Academic Press, chap Global atmospheric modeling using a geodesic grid with an isentropic vertical coordinate., pp 509–538

Scheitlin KN, Elsner JB, Malmstadt JC, Hodges RE, Jagger TH (2010) Toward increased utilization of historical hurricane chronologies. Journal of Geophysical Research-Atmospheres 115, DOI 10.1029/2009JD012424

Smith TM, Reynolds RW, Peterson TC, Lawrimore J (2008) Improvements to NOAAs historical merged LandOcean surface temperature analysis (1880–2006). Journal of Climate 21(10):2283–2296, DOI 10.1175/2007JCLI2100.1

Villarini G, Vecchi GA, Knutson TR, Smith JA (in press) Is the recorded increase in short duration north atlantic tropical storms spurious? Journal of Geophysical Research

Whitaker JS, Compo GP, Wei X, Hamill TM (2004) Reanalysis without radiosondes using ensemble data assimilation. Monthly Weather Review 132(5):1190–1200

Xie L, Yan TZ, Pietrafesa LJ, Morrison JM, Karl T (2005) Climatology and interannual variability of North Atlantic hurricane tracks. Journal of Climate 18(24):5370–5381