EMBRY-RIDDLE
Aeronautical University™
SCHOLARLY COMMONS

Publications

Observed Versus GCM-Generated Local Tropical Cyclone Frequency: Comparisons Using a Spatial Lattice

Sarah Strazzo
Embry-Riddle Aeronautical University, Sarah.Strazzo@erau.edu

Daniel J. Halperin
Embry-Riddle Aeronautical University, halpea17@erau.edu

James Elsner
Florida State University

Tim LaRow
Florida State University

Ming Zhao
Geophysical Fluid Dynamics Laboratory

Follow this and additional works at: https://commons.erau.edu/publication

Part of the Atmospheric Sciences Commons, Climate Commons, Meteorology Commons, Oceanography Commons, and the Statistical Models Commons

SPECIAL
U.S. CLIVAR Hurricanes
COLLECTION

Observed versus GCM-Generated Local Tropical Cyclone Frequency: Comparisons Using a Spatial Lattice

SARAH STRAZZO, JAMES B. ELSNER, TIMOTHY LAROW, AND DANIEL J. HALPERIN

The Florida State University, Tallahassee, Florida

MING ZHAO

Geophysical Fluid Dynamics Laboratory, Princeton, New Jersey

ABSTRACT

Of broad scientific and public interest is the reliability of global climate models (GCMs) to simulate future regional and local tropical cyclone (TC) occurrences. Atmospheric GCMs are now able to generate vortices resembling actual TCs, but questions remain about their fidelity to observed TCs. Here the authors demonstrate a spatial lattice approach for comparing actual with simulated TC occurrences regionally using observed TCs from the International Best Track Archive for Climate Stewardship (IBTrACS) dataset and GCM-generated TCs from the Geophysical Fluid Dynamics Laboratory (GFDL) High Resolution Atmospheric Model (HiRAM) and Florida State University (FSU) Center for Ocean–Atmospheric Prediction Studies (COAPS) model over the common period 1982–2008. Results show that the spatial distribution of TCs generated by the GFDL model compares well with observations globally, although there are areas of over- and underprediction, particularly in parts of the Pacific Ocean. Difference maps using the spatial lattice highlight these discrepancies. Additionally, comparisons focusing on the North Atlantic Ocean basin are made. Results confirm a large area of overprediction by the FSU COAPS model in the south-central portion of the basin. Relevant to projections of future U.S. hurricane activity is the fact that both models underpredict TC activity in the Gulf of Mexico.

1. Introduction

Following recent improvements in model resolution and physics, global climate models (GCMs) are now being employed to study how tropical cyclone (TC) frequency and intensity might change in the future (e.g., Sugi et al. 2002; Knutson and Tuleya 2004; Bengtsson et al. 2007; Gualdi et al. 2008; Knutson et al. 2008; Zhao et al. 2009; Bender et al. 2010; Knutson et al. 2010). TCs can be costly events in terms of loss of life and property. Therefore it is important to understand how these storms may be affected by a warmer climate. GCMs have the potential to be valuable tools in this area of research, but before they can be used with confidence to predict future TC attributes, it is necessary to understand how well they represent historical TCs. Specifically, how well do model-generated TCs match observed TCs with respect to intensity, frequency, and spatial distribution?

After earlier attempts at resolving TC-like vortices in coarse-resolution GCMs (e.g., Manabe et al. 1970; Haarsma et al. 1993; Bengtsson et al. 1995), enhanced computing capabilities and physical parameterizations have resulted in better representations of tropical cyclones in models (e.g., Vitart et al. 1993; Sugi et al. 2002; Oouchi et al. 2006; Bengtsson et al. 2007; Walsh et al. 2007; LaRow et al. 2008; Zhao et al. 2009). Although these finer-resolution GCMs have been able to generate warm core vortices that resemble TCs, the resolution and physics still fall short in some respects. Most notably, the modeled systems are still unable to attain intensities of observed TCs (e.g., Emanuel et al. 2008; LaRow et al. 2008; Zhao and Held 2010). In fact, Chen et al. (2007) used mesoscale models to demonstrate that a grid spacing of ~1 km may be necessary to resolve hurricane eyewall convection and wind maxima. Despite these

Corresponding author address: Sarah Strazzo, Department of Geography, The Florida State University, 113 Collegiate Way, Tallahassee, FL 32306.
E-mail: ses09e@fsu.edu

shortcomings, recent studies reveal some promise in the ability of GCMs to reproduce global TC statistics such as storm counts and seasonal cycle on the basin scale (e.g., Camargo et al. 2005; Zhao et al. 2009). For example, Zhao et al. (2009) ran a global climate model of 50-km horizontal resolution over the period 1981–2005 and found that the correlation of modeled and observed yearly Atlantic Ocean hurricane counts was greater than 0.8, although correlations were lower in the Pacific and Indian Ocean basins.

As resolution and physics continue to improve and new models are developed, intercomparison analyses are needed to provide insight into the model strengths and weaknesses. For example, Camargo et al. (2005) examined genesis location, TC counts, intensity, and storm lifetimes in a statistical analysis of TC-like vortices in three low-resolution GCMs. They found that basin-scale global TC statistics match the observed statistics reasonably well, even for these lower-resolution models. With the suite of higher-resolution models, individual modeling groups have presented TC performance statistics for their specific model (e.g., Bengtsson et al. 2007; Gualdi et al. 2008; LaRow et al. 2008; Zhao et al. 2009). However, there remains a need for a uniform framework that can be used to compare the new suite of high-resolution models with each other and with observations. We present a spatial lattice framework, first introduced for use in TC studies by Elsner et al. (2012), as a novel approach to model intercomparisons. The model comparison we present demonstrates how the spatial lattice approach may facilitate efficient spatial and statistical comparison of model-generated TCs.

Here we employ the spatial lattice to see how well the spatial distribution of actual TCs compares with the distributions generated by two atmospheric GCMs: the Florida State University (FSU) Center for Ocean–Atmospheric Prediction Studies (COAPS) spectral model and the Geophysical Fluid Dynamics Laboratory (GFDL) High-Resolution Atmospheric Model (HiRAM). The methodology applied in this study to compare regional cyclone occurrence between observations and model simulations can be used to compare other models and other storm attributes such as intensity and intensification rate.

The paper is organized as follows: observational and model track data are presented in section 2, followed by an explanation of the spatial lattice methodology in section 3. Section 4 examines the spatial distribution of observed and GFDL model TC occurrences using maps, and section 5 quantifies the comparison using a performance diagram. Section 6 compares the spatial distribution of observed and model TC occurrence over the North Atlantic alone using both the GFDL and FSU

COAPS models. Comparisons are quantified using relative risk ratios. A summary is given in section 7.

2. Data

a. Observational data

Observational data used in this research come from the International Best Track Archive for Climate Stewardship (IBTrACS; available online at http://www.ncdc.noaa.gov/oa/ibtracs/; Knapp et al. 2010). For analysis purposes, the 6-hourly data have been interpolated to hourly intervals using the method outlined in Elsner and Jagger (2013). Although IBTrACS includes more than a century's worth of track data, we subset the data from the period 1982–2008. This is also the time period over which the models were run. Furthermore, because these data were obtained during the satellite era, the reliability of the observations is quite high. Finally, the period of study also includes the transition from the relatively inactive period in the North Atlantic from 1970–94 to the more recent active period after 1995 (e.g., Elsner et al. 2000; Goldenberg et al. 2001). Regardless of the cause, the apparent nonstationary nature of the North Atlantic TC climatology during this period is worth noting.

b. Model data

Model-derived track data are obtained from experiments performed by the Hurricane Working Group of the U.S. Climate Variability and Predictability Research Program (CLIVAR; http://www.usclivar.org/working-groups/hurricane). We use data from two different high-resolution atmospheric (uncoupled) GCMs. As with the observational data, the modeled track data are provided at 6-hourly intervals and have been interpolated to hourly intervals using the same algorithm as used for the observations.

We first use cyclone tracks from the GFDL HiRAM, version 2.2 (Zhao et al. 2009, 2012). The model data come from a control simulation forced with monthly prescribed SSTs and sea ice concentrations for each simulated year from the Hadley Centre Global Sea Ice and Sea Surface Temperature (HadISST) dataset (Rayner et al. 2003). The model features a 50-km horizontal resolution and 32 vertical levels. As described by Zhao et al. (2009), TC-like vortices are detected and tracked using an algorithm similar to that used by Vitart et al. (2003). The algorithm searches for a coinciding (within 2° latitude and longitude) relative vorticity maximum at 850 hPa, a sea level pressure minimum, and a maximum in the 300–500-hPa averaged temperature field. The vortex trajectories are considered TC tracks when the modeled maximum surface winds exceed $15.2\,\mathrm{m\,s^{-1}}$ during at least three (not necessarily consecutive) days (Zhao et al.

TABLE 1. Mean, minimum, and maximum wind speed values ($m\,s^{-1}$) for observations and all three GFDL HiRAM model runs. The last column gives minimum pressure P (hPa). Values shown are only for data exceeding $17\,m\,s^{-1}$.

	Mean	Min	Max	P Min
IBTrACS	31.18	17.00	82.76	878.9
HiRAM (r1)	25.86	17.00	50.82	896.3
HiRAM (r2)	25.44	17.00	52.03	905.4
HiRAM (r3)	26.01	17.00	52.69	897.0

2009). We use track data from three realizations of the HiRAM, referred to here as r1, r2, and r3, which differ only in their initial conditions. The initial conditions in different HiRAM runs are obtained from 1 January of different years from an earlier Atmospheric Model Intercomparison Project (AMIP) run. For these long integrations, the initial conditions are completely irrelevant to the solutions since we do not use the first year results. The purpose is to generate different realizations of the simulation.

We also use cyclone tracks from the FSU COAPS global spectral model (Cocke and LaRow 2000; LaRow et al. 2008). As with the GFDL HiRAM, the FSU COAPS model is uncoupled with the ocean and is forced with prescribed SSTs from the HadISST dataset. The spectral model has 27 vertical levels and a T126 horizontal resolution, which corresponds to roughly 0.94° of latitude. The simulated TC tracks were obtained using a similar algorithm. Because global track data were not available from this model, FSU COAPS tracks are compared with observations and HiRAM tracks for the North Atlantic basin only. Again, we use tracks from three model runs, r1, r2, and r3.

3. The spatial framework

First, a global comparison is implemented using track data from observations and GFDL HiRAM. To set up the spatial framework, we consider only those observations that meet or exceed $17\,m\,s^{-1}$, which corresponds to the 43rd percentile for the global IBTrACS dataset. We also set a $17\,m\,s^{-1}$ minimum threshold for the modeled storms. Relative to observations, the model is unable to produce storms with intensities greater than $50\,m\,s^{-1}$ (category 3 on the Saffir–Simpson scale); however, after the $17\,m\,s^{-1}$ threshold is imposed, the average observed wind speed exceeds the average model wind speeds by less than $6\,m\,s^{-1}$ for all model runs (see Table 1). It should be noted that the models are much more successful in simulating minimum central pressure. The minimum pressures from the GFDL HiRAM track data are 896.3, 905.4, and 897.0 hPa for runs r1, r2, and r3, respectively. These values compare well with the observed minimum

pressure of 878.9 hPa. Therefore, another approach might be to use minimum pressure from the model data to infer wind speeds through some type of pressure–wind relationship and then set the $17\,m\,s^{-1}$ threshold from the inferred wind speeds. However, because we do not have data for the spatial extent of the modeled storms and would introduce additional uncertainty by using a pressure–wind relationship that does not incorporate storm size, we instead simply set the $17\,m\,s^{-1}$ threshold using the raw wind speeds.

For the global spatial analysis, the data locations are first projected onto a planar coordinate system using a Mollweide projection (Snyder 1987). The Mollweide projection is used because we want to examine global distributions of TCs in equal areas. The cost of using an equal-area projection is a sacrifice in the accuracy of angles and shapes, which is significant at the borders of the ellipse.

To establish the spatial lattice, we first define a common grid of equal area hexagons that cover the global tropical and subtropical region to approximately 70°N/S latitude. We choose to use a hexagon lattice because 1) it does not require an areal correction as needed for latitude–longitude grids and 2) relative to rectangles, hexagons are more efficient at capturing the curved nature of TC tracks (Elsner et al. 2012). The use of an areal correction, which would be necessary if we instead relied on latitude–longitude grids, would affect attribute values within the grids. For example, for areally corrected latitude–longitude grids, per-gridcell storm counts will not necessarily be an integer count, thereby making interpretation less intuitive and physical. With an equal-area tessellation, however, counts are preserved as integers. This allows for a more natural and physically meaningful interpretation of the results. We use hexagons of area $7.3 \times 10^5\,km^2$ (slightly larger than the state of Texas) with a diameter, measured from vertex to opposite vertex, of 917 km. The selected area is sufficiently small to capture regional (basin and subbasin) variability. Once the hexagon grid is defined, we populate it with track attribute data from the observational and model datasets. Additional details, justification, and applications for the spatial framework used here can be found in Elsner et al. (2012).

4. Cyclone counts

As we are interested in comparing observed and modeled spatial distributions of storms, we calculate cyclone counts within each hexagon. Cyclone counts represent the number of storms that "passed through" each hexagon in the spatial grid. One count is assigned for each TC that enters a hexagon. For example, a
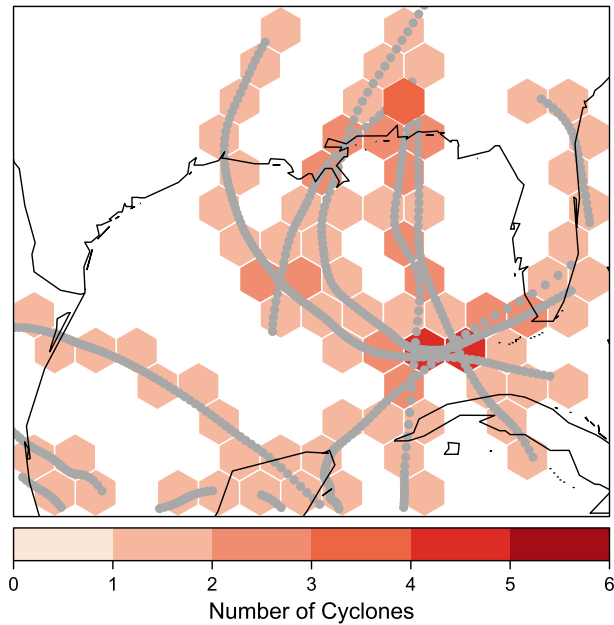
FIG. 1. Observed cyclone tracks (observations in gray) for the Gulf of Mexico in 2005. The colors represent per-hexagon cyclone counts.

FIG. 2. Global per-hexagon cyclone counts for 1982–2008 for (a) observations and (b) GFDL HiRAM (r1).

hexagon centered in the western North Pacific may contain hundreds of storms for the period 1982–2008, while a hexagon in the South Atlantic may contain no storms. Counts are obtained by summing the number of unique storm IDs within each hexagon. Therefore if a TC enters, exits, and then reenters a hexagon, that TC is only counted once within the hexagon. Figure 1 illustrates the framework for the year 2005 in the Gulf of Mexico. We sum the number of cyclones contained in each hexagon and store this information in a polygon data frame for further calculations. This procedure is carried out for both observed and modeled data. It should be noted, however, that the hexagons in Fig. 1 have a smaller area of $1.6 \times 10^5 \, \mathrm{km^2}$ as compared with the hexagons of area $7.3 \times 10^5 \, \mathrm{km^2}$ used in the global analysis that follows.

Using the same spatial lattice to sum TC counts for both observations and model, we can visually compare them with ease. Figure 2 contains a map of cyclone counts for (a) observations and (b) the GFDL HiRAM (r1). Darker reds indicate areas with higher storm frequencies over the 1982–2008 time period. Overall, the maps for observations and the model show a very similar spatial distribution of TCs. In both maps, local maxima in TC counts are present in the eastern and western North Pacific basins as indicated by the darker red hexagons. For observations, 223 out of the 363 hexagons in the lattice have cyclone counts less than 25. Only two hexagons have cyclone counts greater than 200. For the model a total of seven hexagons contain TC counts greater than 200 for HiRAM runs r1, r2, and r3. The
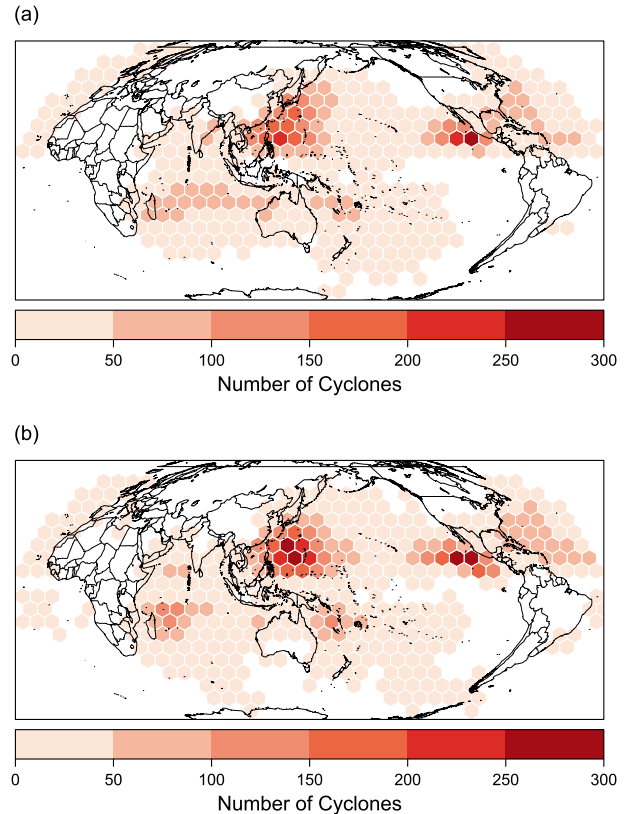
largest observed count is 255, while the largest model counts are 295, 258, and 277 for realizations r1, r2, and r3, respectively. As evident in Fig. 2, there is distortion in the country border lines that is not present when we display the hexagon grid. This is a presentation issue that does not affect the actual analysis. The grid from which the cyclone counts are calculated is properly projected, but the polygons are not distorted as the country borders are when mapped. Furthermore, the grids remain the same for observations and model, so comparison still can be made accurately. However, caution should be exercised when interpreting the precise locations of country borders relative to the hexagons, particularly on the map edges.

Although per-hexagon TC counts greater than 200 may seem high, they occur in particularly active regions of the eastern and western North Pacific basins. The eastern North Pacific is an especially active basin in terms of the number of storms forming over a given unit of area (Gray 1968; Molinari et al. 2000). This is clearly illustrated by the high-count hexagon (from observations) off of the west coast of Mexico. The high-count hexagon in the western North Pacific occurs roughly within the monsoon trough region, as geographically defined in Harr and

Chan (2004). It is estimated that 60%–80% of western North Pacific TCs develop within the monsoon trough (Ritchie and Holland 1999; Harr and Chan 2004), which may help explain the high TC counts in this area. The highest TC count hexagons for the model are similarly located in the eastern and western North Pacific, although the largest counts in the western North Pacific appear to be displaced slightly eastward of the observed maximum.

To facilitate further comparison of the observed and modeled TCs, we also create difference maps (Fig. 3). For cyclone count hexagons, this is accomplished by simply subtracting the per-hexagon modeled storm counts from the per-hexagon observed storm counts. Visually, these results agree well with Zhao et al. (2009). Although the overall distributions of observed and modeled TCs are similar, there is a significant area of overprediction (more modeled than observed cyclones) in the western North Pacific, and a small area of underprediction in the eastern North Pacific (fewer modeled than observed cyclones). There is an additional area of underprediction in the South China Sea. In all three model runs, the areas of strongest overprediction (on the order of 75–125 cyclones) exist in the western North Pacific and in the eastern North Pacific. Interestingly, the model also generates too many storms in the South Atlantic, an area in which observed storms have been exceptionally rare (only one storm was observed here in the 1982–2008 time period). Figures 2 and 3 also demonstrate how the spatial lattice framework may provide additional insight into the range of over/underprediction across space.

5. Metrics of spatial performance

The difference maps allow for a qualitative assessment of model performance in terms of cyclone counts per hexagon. We observe that there are numerous regions in which the model overpredicts the number of TCs (e.g., western North Pacific), while there are other regions in which the model underpredicts (e.g., Gulf of Mexico). Although this is a useful way of assessing subregional model performance in terms of cyclone counts, we are also interested in comparing how well the observed and modeled tracks match spatially over the entire globe. Stated differently, we are interested in knowing whether hexagons that cover areas with observed activity also cover areas with modeled activity. A simple bias calculation provides some insight into this spatial matching aspect of model performance.

To calculate the bias, we compare the set of hexagons in which the observed cyclone count is greater than zero with the set of hexagons in which the modeled cyclone
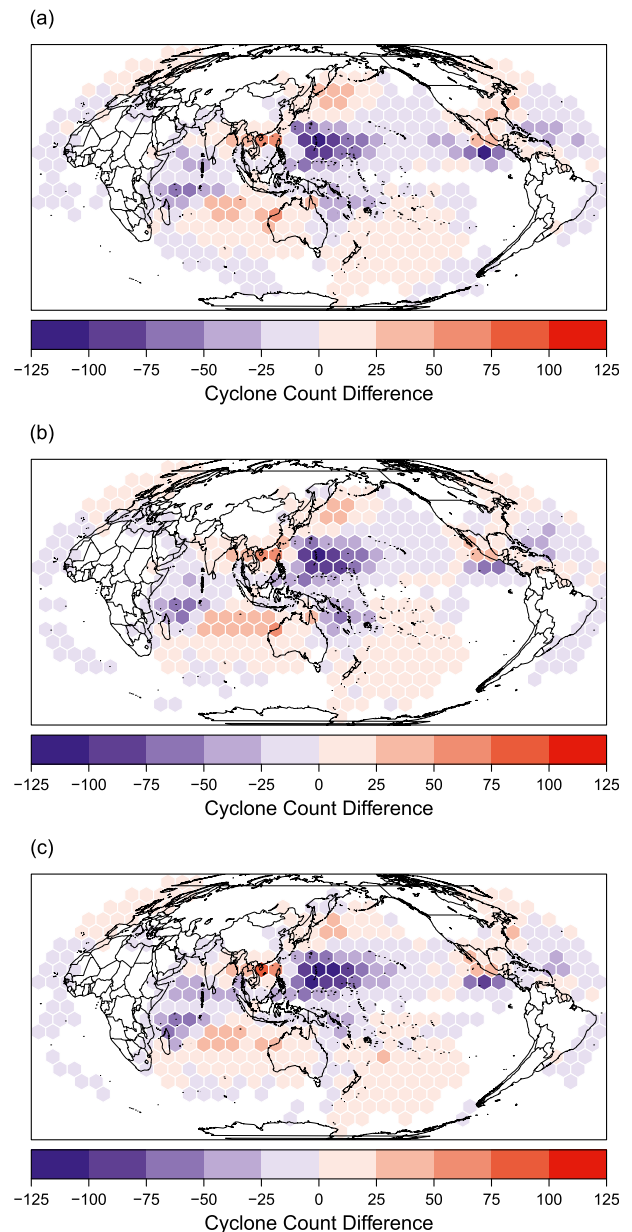

FIG. 3. Difference maps depicting global observed minus global modeled per-hexagon cyclone counts for the GFDL HiRAM (a) r1, (b) r2, and (c) r3.

count is greater than zero. We then calculate the number of hits, false positives, and false negatives from these sets of hexagons. A hit is defined as a hexagon that contains at least one observed and at least one modeled TC; a false positive is defined as a hexagon that contains at least one modeled TC but does not contain an observed TC; and a false negative is defined as a hexagon that contains at least one observed TC but does not contain a modeled TC. We calculate the bias as defined in Roebber (2009):

TABLE 2. Model bias values for the North Atlantic (NA), east Pacific (EP), west Pacific (WP), South Pacific (SP), north Indian (NI), south Indian (SI), South Atlantic (SA), and all basins together (All). The large numbers are median bias values after generating 100 hexagon grids slightly offset from each other in space. The numbers in parentheses represent the 95% confidence intervals. Bias values are provided for varying grid resolutions with per-hexagon areas (10^5 km^2) given in the first column.

Area	NA	EP	WP	SP	NI	SI	SA	All
14.6	0.923	1.13	1.00	1.09	1.05	1.24	8.00	1.13
	(0.874,1.00)	(1.06,1.22)	(0.943,1.09)	(1.03,1.16)	(0.895,1.19)	(1.13,1.34)	(5.33,20.0)	(1.10,1.16)
9.70	0.924	1.16	1.00	1.08	1.04	1.24	8.5	1.12
	(0.860,0.978)	(1.04,1.27)	(0.930,1.09)	(1.01,1.13)	(0.918,1.17)	(1.14,1.36)	(5.41,12.5)	(1.10,1.16)
7.28	0.920	1.15	1.02	1.05	1.00	1.23	9.75	1.11
	(0.864,0.965)	(1.09,1.23)	(0.969,1.07)	(1.00,1.11)	(0.920,1.10)	(1.13,1.34)	(5.24,13.5)	(1.09,1.14)
4.85	0.915	1.15	1.02	1.01	1.00	1.22	9.33	1.10
	(0.868,0.965)	(1.06,1.22)	(0.977,1.07)	(0.979,1.05)	(0.927,1.08)	(1.17,1.29)	(5.10,16.5)	(1.08,1.12)
3.64	0.912	1.15	1.03	0.994	1.00	1.22	10.33	1.08
	(0.874,0.943)	(1.08,1.22)	(0.991,1.06)	(0.959,1.03)	(0.942,1.06)	(1.15,1.28)	(6.40,17.4)	(1.06,1.10)

$$\text{bias} = \frac{H + \text{FP}}{H + \text{FN}}, \qquad (1)$$

where H is the sum of the hits, FP is the sum of the false positives, and FN is the sum of the false negatives for a grid of hexagons populated with observed and model data. For bias values near 1, there is no bias in the model. In this case, the amount of area covered by modeled tracks is the same as the amount of area covered by observed tracks. If the bias is less than 1, the area covered by observed tracks exceeds the area covered by modeled tracks. The model is in a sense underpredicting the spatial extent of the tracks. Conversely, if the bias is greater than 1, the area covered by modeled tracks exceeds the area covered by observed tracks and the model is overpredicting the spatial extent of the tracks.

To obtain an estimate of model bias and its uncertainty relative to our spatial framework, we first generate 100 sets of hexagon lattices. Each lattice is randomly offset very slightly such that the hexagons do not precisely match in space. We then overlay the observed and modeled track data onto each set of hexagons to obtain per-hexagon counts as before. For the model, we use data from the GFDL HiRAM r1 in this section. Hexagons containing at least one observed storm are compared with hexagons containing at least one modeled storm and the bias is then calculated from Eq. (1). These biases are specific to gridcell size. For example, we might expect larger biases from lattices composed of smaller hexagons. To see the effects of grid resolution, we repeat this process for grids with per-hexagon areas of 14.6×10^5, 9.70×10^5, 7.28×10^5 (the original grid area), 4.85×10^5, and 3.64×10^5 km^2. The bias calculations, broken down by ocean basin and gridcell area, are presented with 95% confidence intervals (CIs) in Table 2. The CIs are based on the 2.5th and 97.5th percentiles of

the 100 bias values. Interestingly, varying gridcell sizes in the range presented here does not statistically significantly affect the biases. In general, there is more overprediction than underprediction, with the most consistent exception being the North Atlantic. The overprediction is especially high in the South Atlantic, where the model produced 14 storms over the 1982–2008 period, but only one was observed. In general, there are more false positives than false negatives, although for most basins besides the South Atlantic the bias values do not fall far from 1, indicating good agreement.

From Table 2, we also notice several basins for which the model appears to be unbiased. However, a bias value of 1 is not the perfect summary of model performance. In these cases, the number of false positives may simply balance out the number of false negatives. To gain additional insight to how well the model matches observations within the spatial framework, we also calculate a "critical success index" (CSI) (Roebber 2009). The CSI is expressed as

$$\text{CSI} = \frac{H}{H + \text{FP} + \text{FN}}. \qquad (2)$$

Thus, values near 1 indicate very few false negatives or false positives, or a close match between model and observations, while values closer to zero indicate many false negatives, false positives, or both. Hits (H), false positives (FP), and false negatives (FN) are defined as before. As is done for the biases, we similarly calculate the CSI for 100 lattices and for varying gridcell sizes. These calculations, again broken down by ocean basin and gridcell area, are presented with 95% CIs in Table 3 (expressed as percentages). For most basins besides the South Atlantic and South Pacific, the GFDL HiRAM has relatively high CSI values, generally above 70%. As anticipated, CSI values in the South Atlantic, in which there are numerous false

TABLE 3. Model CSI values (%) for the North Atlantic (NA), east Pacific (EP), west Pacific (WP), South Pacific (SP), north Indian (NI), south Indian (SI), South Atlantic (SA), and all basins together (All). The large numbers are median CSI values after generating 100 hexagon grids slightly offset from each other in space. The numbers in parentheses represent the 95% confidence intervals. CSI values are provided for varying grid resolutions with per-hexagon areas (10^5 km^2) given in the first column.

Area	NA	EP	WP	SP	NI	SI	SA	All
14.6	78.6	84.0	89.3	70.4	84.2	75.7	11.1	75.4
	(73.4,84.5)	(71.4,89.5)	(82.1,94.4)	(64.2,75.8)	(70.0,95.0)	(68.0,81.3)	(5.00,18.8)	(73.4,78.2)
9.70	78.1	80.8	88.6	69.1	82.1	73.6	9.09	74.5
	(73.0,81.2)	(74.8,90.1)	(82.1,95.9)	(64.8,74.7)	(71.4,91.7)	(69.1,79.0)	(4.65,14.7)	(73.4,78.2)
7.28	77.6	81.8	89.3	68.0	79.4	74.2	8.33	74.0
	(72.8,84.0)	(75.0,85.9)	(80.0,94.6)	(64.2,71.9)	(69.0,93.5)	(66.1,79.7)	(4.08,14.3)	(72.8,76.2)
4.85	77.5	80.7	88.7	66.5	78.7	72.2	7.14	73.4
	(73.7,81.6)	(74.3,86.4)	(83.2,92.4)	(63.5,69.4)	(69.8,84.7)	(68.9,76.8)	(6.06,14.3)	(72.0,74.8)
3.64	77.7	79.9	88.2	65.3	75.0	72.0	6.45	72.8
	(73.6,80.4)	(75.3,83.9)	(85.1,91.4)	(62.9,67.0)	(67.5,83.8)	(68.5,75.4)	(5.63,12.0)	(71.9,74.1)

positives, fall below 20% for all grid sizes. Also not entirely unexpected is the slight decrease in CSI that occurs as per-hexagon area decrease; however, these decreases are not statistically significant, as indicated by the overlapping 95% confidence intervals.

Roebber (2009) also introduces a way to visually compare the different metrics of performance in a single plot. We adopt this "performance diagram" to display the bias, CSI, probability of detection (POD), and success rate (SR) for the GFDL HiRAM track data. The probability of detection is expressed as

$$\text{POD} = \frac{H}{H + \text{FN}}. \tag{3}$$

The success rate is expressed as

$$\text{SR} = 1 - \frac{\text{FP}}{H + \text{FP}}. \tag{4}$$

As displayed in Fig. 4, POD is defined along the vertical axis and SR along the horizontal axis. The straight dashed lines in Fig. 4 are lines of equal bias, while the curved solid lines represent lines of equal CSI. Using this diagram, we can visually assess model performance on the basis of four metrics: bias, CSI, POD, and SR. Figure 4 displays these points (with 95% CIs) for the North Atlantic (NA), eastern North Pacific (EP), western North Pacific (WP), South Pacific (SP), north Indian (NI), south Indian (SI), South Atlantic (SA), and for all basins together (ALL). The values plotted represent the median values following the generation of 100 hexagon grids as is done for the bias and CSI in Tables 2 and 3. The values in Fig. 4 are calculated based on grids with a per-hexagon area of 7.28 × 10^5 km^2 (the same area as is used in the maps in Figs. 2 and 3).

Most points in Fig. 4 are clustered in the upper-right portion of the diagram, which indicates a high success

rate and probability of detection. It is also clear from Fig. 4 that most points lie above the "no bias" line (bias = 1), although the western North Pacific and northern Indian Ocean fall very near this line. The North Atlantic is the only basin with a bias less than 1, as mentioned previously. The North Atlantic will be addressed further in the following section. Figure 4 also indicates high CSI
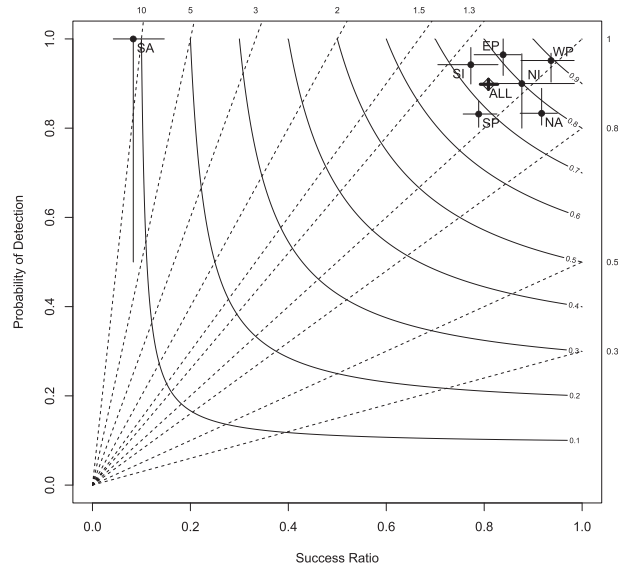

FIG. 4. Performance diagram for the GFDL HiRAM (r1) based on a hexagon grid with a per-hexagon area of 7.28 × 10^5 km^2. The straight dashed lines indicate the bias [as defined in Eq. (1)], the curved solid lines indicate the CSI [as defined in Eq. (2)], and the x and y axes represent the SR [Eq. (4)] and POD [Eq. (3)], respectively. The points represent median values after generating 100 hexagons grids slightly offset from each other in space. Metrics are calculated for the North Atlantic (NA), northern East Pacific (EP), northern west Pacific (WP), South Pacific (SP), north Indian (NI), south Indian (SI), South Atlantic (SA), and all basins together (ALL). The crosshairs indicate the 95% confidence interval for each point.

values for most basins, with the clear exception being the South Atlantic. On the basis of these metrics and our spatial lattice approach, the western North Pacific appears to most closely match observations, with a high probability of detection, a high success rate, a bias near 1, and a CSI close to 0.9. It should be mentioned that, as evident in Fig. 3, the western North Pacific represents a large area of model overprediction in terms of the number of storms present compared to observations. However, in terms of the area covered by TC tracks, the model matches observations well over the western North Pacific.

For the South Atlantic, the performance diagram indicates significant model overprediction based on a high probability of detection, a low success rate, a very low CSI, and a very high bias. This makes sense as only one TC was observed during this period, but 14 TCs were generated by the GFDL HiRAM, r1. Although the GFDL HiRAM produces few TCs over the South Atlantic relative to other basins, for the period 1982–2008 it nevertheless generates far more TCs than observed.

All of the previous calculations are made with a threshold of at least one TC for both model and observations. If we increase the threshold to be at least 15 TCs, for a grid of hexagons with area $7.28 \times 10^5 \, km^2$, the global bias value decreases to 1.01 and the global CSI becomes 83.6% (compared to 1.11 and 74% with a threshold of one TC). For a threshold of 50 TCs, the bias remains at 1.01, but the CSI increases to 90.1%. Finally, for a threshold of at least 200 TCs, the bias still remains near 1.01, but the CSI increases to 98.8%. Therefore, it appears that the areas that generally contain the most observed storms also contain the most modeled storms.

6. Intermodel comparison over the North Atlantic

We next compare the GFDL HiRAM simulated tracks with those generated from the FSU COAPS spectral model. Because global tracks are not yet available for the FSU COAPS model, we focus our comparison on the North Atlantic basin. Once again, we select only observed cyclone points with intensities exceeding $17 \, m \, s^{-1}$, which corresponds to the 33rd percentile of total observed storms over the North Atlantic. As with the HiRAM, a $17 \, m \, s^{-1}$ wind threshold is also set for the FSU COAPS model. The data are projected onto a planar coordinate system using a Lambert conformal conic projection. As was done for the global comparison, a grid of equal-area hexagons is created, this time for the North Atlantic basin. For the North Atlantic comparison, we use hexagons of area $1.91 \times 10^5 \, km^2$, which are much smaller than the hexagons used for the global comparison (slightly larger than the state of Washington). This allows

for a more detailed examination of subregional spatial variability. The hexagons are populated with observations and model data from both the GFDL HiRAM and the FSU COAPS spectral model. Per-hexagon cyclone counts for observations and the first run from each model are shown in Fig. 5.

From Fig. 5 it is apparent that there are some discrepancies between the FSU COAPS model and observations. The FSU COAPS model appears to generate more cyclones over the south-central portion of the North Atlantic. Figure 5b also suggests that most of these model-generated cyclones recurve fairly quickly. This is consistent with LaRow et al. (2008), who attribute this to the model's large-scale steering flow during the first half of the hurricane season. The framework also allows us to directly compare the FSU COAPS model with the HiRAM. In contrast to the FSU COAPS model, which tends to generate too many cyclones over the North Atlantic, the GFDL HiRAM does not generate as many cyclones as are observed. The spatial pattern of the HiRAM cyclones is more consistent with observations, although there are fewer modeled cyclones over the Gulf of Mexico and Caribbean Sea than are observed. This latter point is also true of the FSU COAPS model.

Further comparison of the FSU COAPS model with observations is implemented using relative ratios. We use the ratios to examine the factor by which the modeled cyclone frequency exceeds the observed frequency. This is accomplished by first dividing the number of cyclones in each hexagon by the total number of cyclones for the entire grid. These ratios are calculated for both model and observations. We then divide the model relative ratio by the observed relative ratio to obtain the factor by which modeled cyclone frequency exceeds observed cyclone frequency. The base-two logarithm of these factors, indicated by the color bar, is shown for the FSU model in Fig. 6a. A value greater than zero indicates an overprediction by the model, while a value less than zero indicates an underprediction. The overprediction region is clearly visible in the center of Fig. 6a. It is also apparent that few modeled cyclones are present over the Gulf of Mexico, at higher latitudes, and near the Cape Verde Islands. Figure 6b provides the same information for the GFDL HiRAM. In general, the GFDL HiRAM agrees better with observations, although once again there is a notable lack of model-generated cyclones in the Gulf of Mexico and Caribbean.

Finally, a qualitative comparison between the FSU COAPS and the GFDL HiRAM is made. The results of this model comparison are depicted in Fig. 7. Using the same hexagon lattice as in Figs. 5 and 6, we compare areas in which both models overpredict the number of
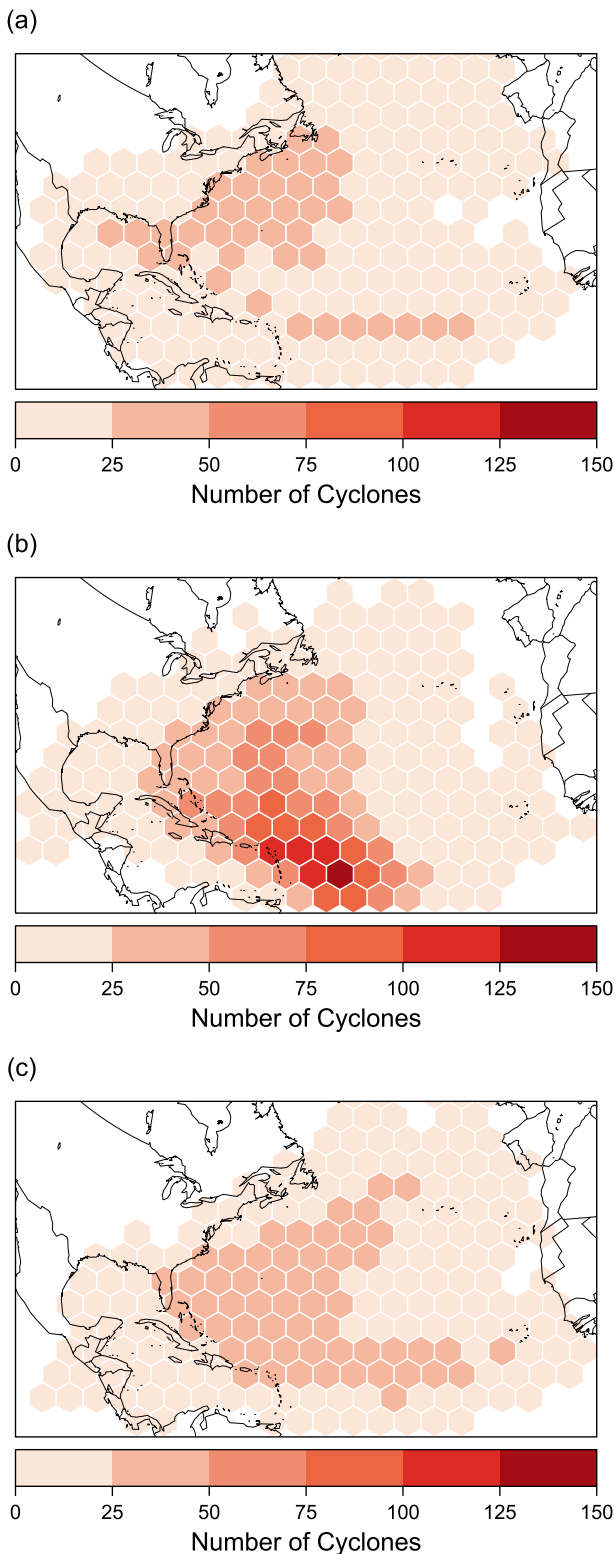
(a)

(b)

(c)

FIG. 5. Per-hexagon cyclone counts in the North Atlantic basin over the 1982–2008 time period for (a) observations, (b) FSU COAPS (r1), and (c) GFDL HiRAM (r1).

(a)

Factor by which model frequency exceeds observed frequency

(b)

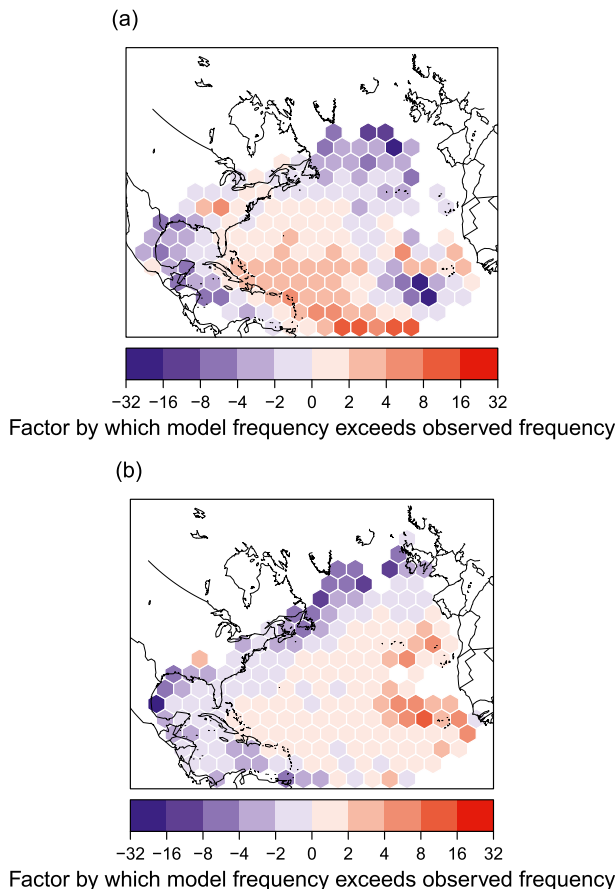Factor by which model frequency exceeds observed frequency

FIG. 6. Hexagons indicate the factor by which model storm frequency exceeds observed storm frequency for (a) FSU COAPS and (b) GFDL HiRAM. Values greater than 0 (pinks and reds) indicate that modeled storm frequency exceeds observed storm frequency, while values less than 0 (blues) indicate that observed storm frequency exceeds modeled storm frequency.

TCs relative to observations (shown in red), the areas in which the FSU model overpredicts but the GFDL underpredicts (magenta), the areas in which the GFDL overpredicts but the FSU underpredicts (cyan), and finally the areas in which both models underpredict (blue). It is clear from this map that both models underpredict over both the Caribbean and Gulf of Mexico. This is perhaps the result of a lack of model genesis over this subregion, premature recurving of storms generated farther east, or both. In the FSU COAPS model, for example, LaRow et al. (2008) note that a break in the ridge over the central Atlantic during the peak of the hurricane season allows more modeled storms to recurve early rather than continue westward.

The possibility of a lack of model genesis over the Gulf of Mexico is also addressed. Figure 8 displays counts for per-hexagon genesis points for observations (Fig. 8a), FSU COAPS (Fig. 8b), and GFDL HiRAM (Fig. 8c). To
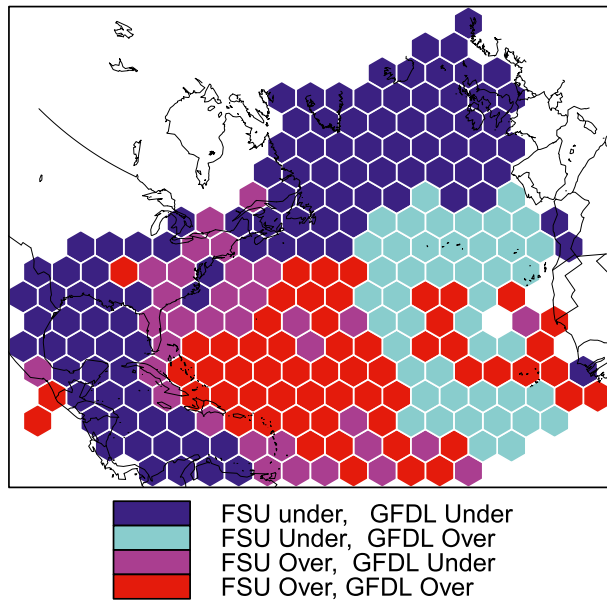
FIG. 7. Hexagons indicate subregions in which there was over-prediction by both the FSU and GFDL models (red), over-prediction by the FSU model and underprediction by the GFDL (magenta), underprediction by the FSU model and overprediction by the GFDL (cyan), or underprediction by both models (blue).

obtain the genesis points, we define "genesis" as the first record for each individual storm in the IBTrACS and model datasets. As expected, hexagons with the most observed genesis points are located off the west coast of Africa, throughout the main development region (MDR), and also over the Caribbean and Gulf of Mexico. In contrast to observations, nearly all genesis in the FSU COAPS model occurs over the south-central portion of the basin, well west of Africa and east of the Caribbean and Gulf of Mexico. In fact, the highest per-hexagon genesis count for the FSU COAPS model is 50 storms, significantly higher than the maxima of 13 and 15 for observations and the GFDL, respectively. Very few storms generated by the FSU COAPS model form over the Gulf of Mexico or Caribbean. The distribution of genesis points from the GFDL HiRAM more closely matches observations; however, there is still a notice-able lack of model genesis over the Gulf of Mexico and Caribbean by this model as well. For the GFDL HiRAM, the small number of modeled TCs developing over the Gulf of Mexico and Caribbean may be a result of large model wind shear anomalies, although another possible cause is the general lack of simulated convec-tive activity over this region. This is also true of the FSU COAPS model, which generates higher than observed wind shear over the Gulf of Mexico for the August–October period. The FSU COAPS model also displays a dry precipitation bias in this region.
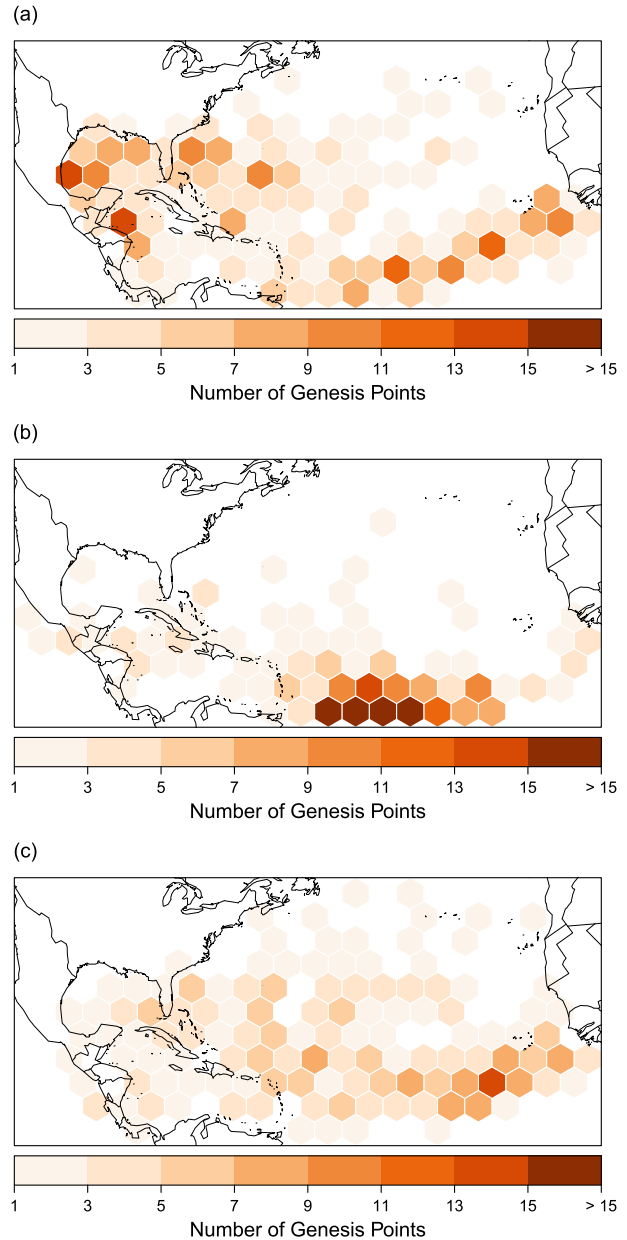

FIG. 8. Per-hexagon counts of genesis points for (a) observations, (b) FSU COAPS, and (c) GFDL HiRAM. "Genesis" is defined as the first record for each storm listed in the IBTrACS and model datasets. The darkest shading represents genesis counts greater than 15.

In addition to the underprediction over the Gulf of Mexico and Caribbean, the area of overprediction by the FSU COAPS model is evident in the red and magenta hexagons of Fig. 7. It is also interesting to note that the area of GFDL HiRAM overprediction extends farther north and east across the basin. Although Fig. 7 provides a qualitative assessment of model over- and/or underprediction, the magnitude of the model discrep-ancy is not apparent.

7. Summary

GCMs are now routinely employed to study how TC frequency may change with a warmer climate. However, before confidence can be placed in future cyclone scenarios, it is necessary to understand how well they reproduce the historical spatial climatology. Using a methodology based on the spatial tessellation of Elsner et al. (2012), this study puts forward a spatial lattice approach to quantitatively compare regional TC activity.

Global and regional comparisons are made between actual and simulated TC occurrences using actual TCs from the IBTrACS dataset and GCM-generated TCs from the GFDL HiRAM and FSU COAPS models over the common period 1982–2008. Globally results show that although there are some areas of over- and underprediction, the spatial distribution of TCs generated by the GFDL HiRAM compare well with observations. Difference maps using the spatial lattice highlight the areas in which the model disagrees with observations. The primary mismatch areas are found in the Pacific. Several quantitative metrics of model success are used to examine the ability of the GFDL HiRAM to accurately capture the spatial extent of TC tracks globally and regionally. Again, the model performs fairly well overall, with the primary problem area being the South Atlantic. Additionally, comparisons focusing on the North Atlantic basin are made using both models. Results confirm a large area of overprediction by the FSU COAPS model over the south-central portion of the basin, and a large area of underprediction by both models over the Gulf of Mexico and Caribbean. The underprediction is particularly relevant to projections of future U.S. hurricane activity.

As the data become publicly available, this method can be applied to provide a comprehensive model comparison using all model datasets from phase 5 of the Coupled Model Intercomparison Project (CMIP5). All the code used to generate the results of this paper is available online (http://rpubs.com/sestrazz/4591).

REFERENCES

Bender, M. A., T. R. Knutson, R. E. Tuleya, J. J. Sirutis, G. A. Vecchi, S. T. Garner, and I. Held, 2010: Modeled impact of anthropogenic warming on the frequency of intense Atlantic hurricanes. *Science,* **327,** 454–458.

Bengtsson, L., M. Botzet, and M. Esch, 1995: Hurricane-type vortices in a general circulation model. *Tellus,* **47A,** 175–196.

——, K. I. Hodges, and M. Esch, 2007: Tropical cyclones in a T159 resolution global climate model: Comparison with observations and re-analyses. *Tellus,* **59A,** 396–416.

Camargo, S. J., A. G. Barnston, and S. E. Zebiak, 2005: A statistical assessment of tropical cyclone activity in atmospheric general circulation models. *Tellus,* **57A,** 589–604.

Chen, S. S., W. Zhao, M. A. Donelan, J. F. Price, and E. J. Walsh, 2007: The CBLAST-hurricane program and the next-generation fully coupled atmosphere–wave–ocean models for hurricane research and prediction. *Bull. Amer. Meteor. Soc.,* **88,** 311–317.

Cocke, S. D., and T. E. LaRow, 2000: Seasonal predictions using a regional spectral model embedded within a coupled ocean–atmosphere model. *Mon. Wea. Rev.,* **128,** 689–708.

Elsner, J. B., and T. H. Jagger, 2013: *Hurricane Climatology: A Modern Statistical Guide Using R.* Oxford University Press, 390 pp.

——, ——, and X.-F. Niu, 2000: Changes in the rates of North Atlantic major hurricane activity during the 20th century. *Geophys. Res. Lett.,* **27,** 1743–1746.

——, R. E. Hodges, and T. H. Jagger, 2012: Spatial grids for hurricane climate research. *Climate Dyn.,* **39,** 21–36.

Emanuel, K. A., R. Sundararajan, and J. Williams, 2008: Hurricanes and global warming: Results from downscaling IPCC AR4 simulations. *Bull. Amer. Meteor. Soc.,* **89,** 347–367.

Goldenberg, S. B., C. W. Landsea, A. M. Mestas-Nuñez, and W. M. Gray, 2001: The recent increase in Atlantic hurricane activity: Causes and implications. *Science,* **293,** 474–479.

Gray, W. M., 1968: Global view of the origin of tropical disturbances and storms. *Mon. Wea. Rev.,* **96,** 669–700.

Gualdi, S., E. Soccimarro, and A. Navarra, 2008: Changes in tropical cyclone activity due to global warming: Results from a high-resolution coupled general circulation model. *J. Climate,* **21,** 5204–5228.

Haarsma, R. J., J. F. B. Mitchell, and C. A. Senior, 1993: Tropical disturbances in a GCM. *Climate Dyn.,* **8,** 247–257.

Harr, P. A., and J. C. L. Chan, 2004: Monsoon impacts on tropical cyclones variability. *Proc. Third Int. Workshop on Monsoons,* Hangzhou, China, WMO, 512–542.

Knapp, K. R., M. C. Kruk, D. H. Levinson, H. J. Diamond, and C. J. Neumann, 2010: The International Best Track Archive for Climate Stewardship (IBTrACS): Unifying tropical cyclone data. *Bull. Amer. Meteor. Soc.,* **91,** 363–376.

Knutson, T. R., and R. E. Tuleya, 2004: Impact of CO_2-induced warming on simulated hurricane intensity and precipitation: Sensitivity to the choice of climate model and convective parameterization. *J. Climate,* **17,** 3477–3495.

——, J. J. Sirutis, S. T. Garner, G. A. Vecchi, and I. M. Held, 2008: Simulated reduction in Atlantic hurricane frequency under twenty-first-century warming conditions. *Nat. Geosci.,* **1,** 359–364.

——, and Coauthors, 2010: Tropical cyclones and climate change. *Nat. Geosci.,* **3,** 157–163.

LaRow, T. E., Y. K. Lim, D. W. Shin, E. P. Chassignet, and S. Cocke, 2008: Atlantic basin seasonal hurricane simulations. *J. Climate,* **21,** 3191–3206.

Manabe, S., J. L. Holloway, and H. M. Stone, 1970: Tropical circulation in a time-integration of a global model of the atmosphere. *J. Atmos. Sci.,* **27,** 580–613.

Molinari, J., D. Vollaro, S. Skubis, and M. Dickinson, 2000: Origins and mechanisms of eastern Pacific tropical cyclogenesis: A case study. *Mon. Wea. Rev.,* **128,** 125–139.

Oouchi, K., J. Yoshimura, H. Yoshimura, R. Mizuta, and A. Noda, 2006: Tropical cyclone climatology in a global-warming climate as simulated in a 20 km-mesh global atmospheric model: Frequency and wind intensity analyses. *J. Meteor. Soc. Japan,* **84,** 259–276.

Rayner, N. A., D. E. Parker, E. B. Horton, C. K. Folland, L. V. Alexander, D. P. Powell, E. C. Kent, and A. Kaplan, 2003: Global analyses of sea surface temperature, sea ice, and night marine air temperature since the late nineteenth century. *J. Geophys. Res.,* **108,** 4407, doi:10.1029/2002JD002670.

Ritchie, E. A., and G. J. Holland, 1999: Large-scale patterns associated with tropical cyclogenesis in the western Pacific. *Mon. Wea. Rev.,* **127,** 2027–2043.

Roebber, P. J., 2009: Visualizing multiple measures of forecast quality. *Wea. Forecasting,* **24,** 601–608.

Snyder, J. P., 1987: Map projections—A working manual. U.S. Geological Survey Prof. Paper 1395, 383 pp.

Sugi, M., A. Noda, and N. Sato, 2002: Influence of the global warming on tropical cyclone climatology: An experiment with the JMA global model. *J. Meteor. Soc. Japan,* **80,** 249–272.

Vitart, F., J. L. Anderson, and W. F. Stern, 1993: Impact of large-scale circulation on tropical storm frequency, intensity, and location, simulated by an ensemble of GCM integrations. *J. Climate,* **12,** 3237–3254.

——, D. Anderson, and T. Stockdale, 2003: Seasonal forecasting of tropical cyclone landfall over Mozambique. *J. Climate,* **16,** 3932–3945.

Walsh, K. J. E., M. Fiorino, C. W. Landsea, and K. L. McInnes, 2007: Objectively determined resolution-dependent threshold criteria for the detection of tropical cyclones in climate models and reanalyses. *J. Climate,* **20,** 2307–2314.

Zhao, M., and I. M. Held, 2010: An analysis of the effect of global warming on the intensity of Atlantic hurricanes using a GCM with statistical refinement. *J. Climate,* **23,** 6382–6393.

——, ——, S. J. Lin, and G. A. Vecchi, 2009: Simulations of global hurricane climatology, interannual variability, and response to global warming using a 50-km resolution GCM. *J. Climate,* **22,** 6653–6678.

——, ——, and ——, 2012: Some counterintuitive dependencies of tropical cyclone frequency on parameters in a GCM. *J. Atmos. Sci.,* **69,** 2272–2283.