A Poisson regression index for tropical cyclone genesis and the role of large-scale vorticity in genesis

MICHAEL K. TIPPETT*

International Research Institute for Climate and Society, Columbia University, Palisades, New York

SUZANA J. CAMARGO

Lamont-Doherty Earth Observatory, Columbia University, Palisades, New York

ADAM H. SOBEL

Department of Applied Physics and Applied Mathematics, Department of Earth and Environmental Sciences, Columbia University, New York, New York

**Corresponding author address:* M. K. Tippett, International Research Institute for Climate and Society, The Earth Institute of Columbia University, Lamont Campus / 61 Route 9W, Palisades New York 10964, USA.

E-mail: tippett@iri.columbia.edu

ABSTRACT

A Poisson regression between the observed climatology of tropical cyclogenesis (TCG) and large-scale climate variables is used to construct a TCG index. The regression methodology is objective and provides a framework for the selection of the climate variables in the index. Broadly following earlier work, four climate variables appear in the index: low-level absolute vorticity, relative humidity, relative sea surface temperature (SST) and vertical shear. Several variants in the choice of predictors are explored, including relative SST vs. potential intensity and satellite-based column integrated relative humidity vs. reanalysis relative humidity at a single level; these choices make modest differences in the performance of the index. The feature of the new index which leads to the greatest improvement is a functional dependence on low-level absolute vorticity that causes the index response to absolute vorticity to saturate when absolute vorticity exceeds a threshold. This feature reduces some biases of the index and improves the fidelity of its spatial distribution to the observed climatology of genesis. Physically, this result suggests that once low-level environmental vorticity reaches a sufficiently large value, other factors become rate-limiting so that further increases in vorticity (at least on a monthly mean basis) do not increase the probability of genesis.

Although the index is fit to climatological data, it reproduces some aspects of interannual variability when applied to interannually varying data. Overall the new index compares positively to the genesis potential index (GPI) whose derivation, computation and analysis is more complex in part due to its dependence on potential intensity.

1. Introduction

We are interested in the relationship between the statistical distribution of tropical cyclone genesis (TCG) and the large-scale climate. If the climate changes, either due to natural or anthropogenic causes, will there be more or fewer tropical cyclones in a given basin? Will their spatial distribution within the basin change? If it does change, what climate factors are most important in producing that change and why?

At present we lack a solid theoretical foundation with which to answer these questions from first principles. Numerical models are becoming able to provide plausible answers as available computational power now permits the use of global high-resolution models which simulate both the global climate and tropical cyclones with some fidelity (e.g., Oouchi et al. 2006; Bengtsson et al. 2007; Gualdi et al. 2008; Zhao et al. 2009). These models are expensive, however, and are subject to all the normal limitations of numerical models: they may be biased, and simulations with them do not automatically provide understanding. Empirical study of the problem using the observational record remains relevant.

Gray (1979) developed an empirical "index" for tropical cyclogenesis. Gray's index, and those which have been developed later following Gray's example, are functions of a set of predictors–physical fields to which genesis is believed to be sensitive, usually computed from large-scale data and often averaged over a month or some comparable duration–weighted in such a way that larger values of the index are indicative of a greater probability of genesis. Later investigators, following the same basic approach, have tried to improve on Gray's index by changing either the predictors or the functional dependence of the index on them (e.g., DeMaria et al. 2001; Royer et al. 1998; Emanuel and Nolan 2004; Camargo et al. 2007a; Sall et al. 2006; Bye and Keay 2008; Kotal et al.

2009; Murakami and Wang 2010).

The present study also follows the same basic approach, with some incremental improvements. We aim to improve both the performance of the index and the degree to which its derivation can be understood and reproduced. To motivate this work we describe some limitations of the Emanuel and Nolan GPI, (as described in more detail by Camargo et al. 2007a) which we take to be more or less representative of the state of the art. The GPI has been applied widely and successfully to study variations of genesis frequency on various time scales in reanalysis and models (e.g, Camargo et al. 2007a,b; Vecchi and Soden 2007b; Nolan et al. 2007; Camargo et al. 2009; Lyon and Camargo 2009; Yokoi et al. 2009; Yokoi and Takayuba 2009). However, the GPI has the following limitations:

i. Its derivation was partly subjective, and thus cannot be easily reproduced.

ii. One of its thermodynamical predictors, potential intensity (PI), is a highly derived quantity whose computation requires use of a sophisticated algorithm. It is not clear whether this degree of technical difficulty and theoretical complexity is necessary, or whether equal performance can be obtained with a simpler predictor.

iii. The choice of relative humidity at a single level as the other thermodynamic predictor is not precisely justified, and in practice generally requires use of assimilated humidity fields which are not strongly constrained by observations.

iv. The GPI itself compared to the observed genesis climatology has some systematic biases. For example: in the seasons when no tropical cyclones are observed, it continues to predict a non-negligible probability of genesis; and in some regions of particular interest, such as the tropical Atlantic Main Development Region, the GPI under-predicts the rate of genesis during the peak season.

Of these limitations, the first is perhaps the most significant. The importance of the problem warrants development of an index by a process which is clear, objective, and reproducible. The procedure should make apparent the consequences of the choices made, so that the procedure can be easily varied and adapted due to the requirements of some particular application, to make use of new observations of the predictors (or different predictors), or for other reasons unforeseen.

The second and third limitations are more minor, but still worthy of consideration. PI was chosen by Emanuel and Nolan (2004) to replace the thermodynamic parameter used by Gray. Gray's thermodynamic parameter is proportional to the difference between upper ocean heat content (sometimes replaced in later work by sea-surface temperature (SST)) and a fixed threshold value, below which genesis is assumed impossible (Gray 1979). While deep convection does seem to be very roughly parameterizable as depending on SST above some threshold (e.g., Back and Bretherton 2009), our current understanding is that this threshold should not be fixed but should vary with the mean climate, whether due to anthropogenic or natural causes. PI depends on the mean climate in a way that may be plausibly assumed to capture this dependence, and this is an improvement on Gray's SST predictor for studying the influence of large-scale climate variability and change on genesis frequency.

However, the computation of PI involves a complex algorithm, which makes it difficult both to compute the index and to understand its behavior. Moreover, the use of PI does not add a great deal of theoretical justification, since PI is a theoretical prediction of the maximum tropical cyclone intensity rather than the likelihood of genesis (at which point, by definition, tropical cyclone intensity is at a minimum). It is not clear that simpler predictors could not provide comparable performance. In particular, relative SST–the difference between the local SST and the mean tropical SST–has been shown to be highly correlated with PI (Vecchi and Soden 2007a; Swanson 2008) as can be explained by straightforward physical arguments (Sobel et al. 2002; Ramsay and Sobel 2010). Relative SST is similar to Gray's original SST parameter, but differs in being a linear function of SST (no Heaviside function) and in allowing for change in the mean climate by using tropical mean SST in place of a fixed value.

There is little doubt that free-tropospheric relative humidity is a factor in tropical cyclogenesis (e.g., Gray 1979; Emanuel 1989; Cheung 2004) but the relative influence of different parts of the vertical profile of relative humidity is not precisely known. Therefore the choice of level or levels to use as predictors in an index is somewhat arbitrary. As microwave satellite retrievals of column-integrated water vapor are available, it seems reasonable to consider this quantity as a predictor in place of reanalysis products. Given the possible biases in either satellite retrievals or reanalysis products, it should be noted that the regression can implicitly correct systematic errors in its inputs.

In this study, we address the first three limitations, and examine the extent to which choices we make in the development of the index influence its performance (the fourth). It turns out that the change which leads to the most significant improvement in performance is not related to any of the first three issues above, but rather involves the vorticity parameter. This result in turn has implications about the physics of genesis. While the presence of environmental vorticity is required for TCG, we find that the sensitivity of TCG to absolute vorticity is nonlinear. When the absolute vorticity exceeds a threshold, further increase does not increase the likelihood of TCG. Inclusion of this functional dependence in the index significantly improves the index performance.

The paper is organized as follows. Section 2 addresses the data used and the Poisson regression methodology used to construct the index. Section 3 details the construction of the index. Section 4 examines the properties of the index, including: spatial distribution, basin-scale quantities, dependence on climate variables, seasonal cycle and interannual variability. Summary and conclusions are given in Section 5.

2. Data and methodology

a. Data

All data are represented on a $2.5° \times 2.5°$ longitude-latitude grid extending from 60°S to 60°N. Values of 850 hPa absolute vorticity, 600 hPa relative humidity and vertical shear between the 850 hPa and 200 hPa levels come from the monthly mean values of the National Centers for Environmental Prediction (NCEP)-National Center for Atmospheric Research (NCAR) Reanalysis (Kalnay and Coauthors 1996; Kistler et al. 2001) and the 40-year European Centre for Medium-Range Weather Forecasts (ECMWF) Re-Analysis (ERA-40) (Uppala et al. 2005) data sets. Climatological means of the variables for both datasets were computed using the common 40-year period 1961-2000.

The column-integrated relative humidity was calculated following the procedure developed in Bretherton et al. (2004). We obtained the retrievals of column-integrated liquid water W from the Remote Sensing Systems Inc. (see `http://www.remss.com`) for all available Special Sensor Microwave Imager [(SSMI) satellites F08, F10, F11, F12, F14, and F15] in the period 1987-2008. Details of the SSMI retrieval algorithms is given in Wentz and Spencer (1998). The data are provided on a $0.25° \times 0.25°$ grid and are suitable for use over the ocean only. First the daily average over each ocean grid point was calculated based on all valid data, then these averaged daily data were rescaled to a $2.5° \times 2.5°$ grid for the region 60°S to 60°N. Following Bretherton et al. (2004), we calculated the daily averaged saturation water vapor path W_*, using the daily temperature data and surface pressure from the ERA-40 and NCEP reanalysis. The saturation specific humidity was calculated at each pressure level and grid point, and then vertically integrated for each day of the common period when each reanalysis and SSMI data were available (NCEP: 1987-2008, ERA: 1987- August 2002). The daily column-relative humidity is then defined as the ratio W/W_*. Then monthly means are obtained, from which climatological values can be calculated.

The relative SST is defined as the SST at each grid point minus the mean SST of the 20°S-20°N region (Vecchi and Soden 2007a; Vecchi et al. 2008). The SST product used was version 2 of the National Oceanic and Atmospheric Administration (NOAA) National Climatic Data Center (NCDC) extended reconstruction sea surface temperature (ERSST2; Smith and Reynolds 2004).

The potential intensity (PI) was obtained from monthly means of ERSST2, sea level pressure, and vertical profiles of atmospheric temperature and humidity for both NCEP and ERA reanalysis datasets. The algorithm developed by Kerry Emanuel is a generalization of the procedure described in Emanuel (1995) taking into account dissipative heating (Bister and Emanuel 1998, 2002a,b).

The genesis potential index (GPI) was developed by Emanuel and Nolan (2004) and discussed in detail in Camargo et al. (2007a) and was also used in Camargo et al. (2007b); Nolan et al. (2007); Vecchi and Soden (2007a); Camargo et al. (2009). The genesis potential index is defined as

$$\text{GPI} = \left|10^5\eta\right|^{3/2} \left(\frac{\mathcal{H}}{50}\right)^3 \left(\frac{\text{PI}}{70}\right)^3 (1+0.1V)^{-2}, \tag{1}$$

where η is the absolute vorticity at 850hPa in s^{-1}, $\mathcal{H}$ is the relative humidity at 600hPa in percent, PI is the potential intensity in ms^{-1}, and V is the magnitude of the vertical wind shear between 850hPa and 200hPa in ms^{-1}.

The tropical cyclone genesis data used were obtained from various best-track datasets for the period 1961-2000. The best-track dataset for the Atlantic and eastern North Pacific is from the National Hurricane Center. The Joint Typhoon Warning Center best-track datasets were used for the western North Pacific, North Indian Ocean and southern hemisphere.

b. Poisson regression

Poisson regression is appropriate for the modeling of count data such as TCG occurrence (Solow and Nicholls 1990; Elsner and Schmertmann 1993; McDonnell and Holbrook 2004; Mestre and Hallegatte 2009; Villarini et al. 2010). A random variable N has a Poisson distribution with expected value μ if N takes on the values $n = 1, 2, 3, \ldots$ with probability

$$P(N = n) = \frac{e^{-\mu}\mu^n}{n!}\,. \tag{2}$$

Here, N is the number of TCG events during a 40-year climatological period. Our goal is to predict the expected value μ from a vector $\mathbf{x}$ of climate variables. A model in which the expected value μ depends linearly on the climate variables $\mathbf{x}$ is unsatisfactory since negative values of μ may result. A solution is to use a log-linear model where $\log \mu$ is linearly related to $\mathbf{x}$, that is,

$$\log \mu = \mathbf{b}^T\mathbf{x} \tag{3}$$

where $\mathbf{b}$ is a vector of coefficients, or equivalently,

$$\mu = \exp(\mathbf{b}^T\mathbf{x})\,. \tag{4}$$

A constant term (intercept) is included in the model by taking one of the elements of $\mathbf{x}$ to be unity. This model, where the number N of TCG events has a Poisson distribution and the logarithm of its expected value is a linear combination of predictors, is a Poisson regression model, a special case of a generalized linear model.

The climate variables and the observed number of TCG events are defined on a $2.5^\circ \times 2.5^\circ$ latitude-longitude grid. To account for the differing area associated with grid points at different latitudes we modify (4) to become

$$\mu = \exp(\mathbf{b}^T\mathbf{x} + \log\cos\phi)\,, \tag{5}$$

where ϕ is the latitude. The offset term $\log\cos\phi$ is a predictor with coefficient one and serves to make the units of $\exp(\mathbf{b}^T\mathbf{x})$ be number of TCG events per area.

The log-likelihood L of k independent observations $N_1, N_2, \ldots, N_k$ drawn from Poisson distributions with means $\mu_1, \mu_2, \ldots, \mu_k$ is from (2)

$$L = \sum_{i=1}^{k} N_i \log\mu_i - \mu_i - \log(N_i!)\,. \tag{6}$$

The mean μ_i depends on the associated climate variables $\mathbf{x}_i$ and the coefficients $\mathbf{b}$ through the relation in (5). Therefore, for specified observations N_i and climate variables $\mathbf{x}_i$, the log-likelihood L is a function only of the unknown coefficients $\mathbf{b}$. The coefficients are found by maximizing the log-likelihood L defined in (6). Here we fit a single Poisson regression model to climatological data for all ocean grid points and months of the year; the subscript i indexes grid points and months. Moreover, we include both the NCEP and ERA data in (6), thus allowing the estimation of a single Poisson regression model relating the expected number of TCG events with climate variables.

The maximized log-likelihood is a measure of how well the model fits the data. However, since the maximized log-likelihood is the result of an optimization, it is positively biased and this bias increases as the number of predictors increases. This bias is reflected in the fact that the maximized log-likelihood always improves as the number of predictors increases, regardless of whether the additional predictors would prove useful on independent data. The Akaike information criterion (AIC) corrects for that bias and attempts to avoid selection of useless predictors and overfitting (Akaike 1973). The AIC is defined

$$\mathrm{AIC} = -2L + 2p \tag{7}$$

where p is the number of parameters in the model. The AIC is oriented so that models with lower AIC are considered superior. The first term rewards model fit while the second term penalizes models with many parameters. However, AIC is a function of the data and therefore random, and as such, should only be used as a guide in predictor selection selection.

3. Construction of a TCG index

Our goal is to construct an index which reflects the dependence of TCG on large-scale climate variables. Necessary conditions for TCG include sufficient environmental vorticity, humidity, ocean thermal energy and lack of vertical shear. There are several questions regarding how to include these factors in a TCG index. The ones that we consider here are the following:

- Is reanalysis relative humidity adequate given that there are relatively few humidity measurement in the oceanic troposphere (Kistler et al. 2001)? Is there any benefit from using column-integrated relative humidity from satellite microwave retrievals?
- Should relative SST, or PI be used to represent the availability of ocean thermal energy?
- To what extent is the dependence of the number of TCG events on the climate variables log-linear?

We first broadly address these questions in the framework of predictor selection, using the AIC to assess how well the index fits the observations. Later, in Section 4, we examine these questions in terms of basin-integrated quantities and spatial distributions.

Initially, we consider the four climate variables: absolute vorticity, reanalysis mid-level relative humidity, relative SST and vertical shear. The simplest Poisson regression model assumes no interactions between the predictor variables, i.e., no powers or products of the predictors are included in the model, and has the form

$$\mu = \exp\left(b + b_\eta \eta + b_{\mathcal{H}} \mathcal{H} + b_T T + b_V V + \log\cos\phi\right) \tag{8}$$

where μ is the expected number of tropical cyclone genesis events in 40 years, and η, $\mathcal{H}$, T and V are, respectively, the absolute vorticity at 850 hPa in $10^5 s^{-1}$, the relative humidity at 600 hPa in percent, relative SST in °C and vertical shear between the 850 hPa and 200 hPa levels in ms^{-1}; b is the constant (intercept) term; we adopt the convention that the coefficient subscript indicates the quantity it multiplies; in the notation of (5), $\mathbf{b} = [b, b_\eta, b_{\mathcal{H}}, b_T, b_V]$ and $\mathbf{x} = [1, \eta, \mathcal{H}, T, V]$

Maximizing the likelihood (6) of the observed number of TCG events given the NCEP and ERA climatological data leads to the coefficient values shown in the first line of Table 1. The form of the Poisson regression model means that the coefficients can be directly interpreted as sensitivities. Specifically, for a small change $\delta\mathbf{x}$ in the climate variables, the change $\delta\mu$ in the mean number of TCG events is

$$\frac{\delta\mu}{\mu} \approx \mathbf{b}^T \delta\mathbf{x}\,. \tag{9}$$

That is, for a 0.01 unit change in one of the climate variables, the corresponding coefficient is the percent change in μ. For instance, an increase of 1 cm s^{-1} in vertical shear reduces the expected number of TCG events by about 0.15% in this model.

Addressing the first question regarding the choice of humidity variable, we find that using the SSMI column-integrated relative humidity rather than the reanalysis relative humidity gives a lower AIC value (second line of Table 1), indicating a better fit to observations. However, fitting the NCEP and ERA data separately reveals that using the SSMI column-integrated relative humidity improves the fit for NCEP data but not for ERA data. Figure 1 shows that overall the ERA relative humidity has a stronger seasonal cycle than NCEP in the northern hemisphere. NCEP relative humidity differs considerably from ERA in the southern hemisphere during austral winter. The difference between the NCEP and ERA relative humidity do not facilitate the use of a single Poisson regression model. The hemisphere and basin averaged column-integrated relative humidities computed using SSMI water vapor path divided by saturation values derived from either NCEP and ERA temperatures are more similar to each other (not shown).

SST and PI are both variables that can be used to quantify the availability of ocean heat for TCG and either could conceivably be used in the index in place of relative SST. For fitting the spatial distribution of genesis probability from the present mean climatology, SST contains nearly the same information as relative SST. As the climate varies (due to either anthropogenic or natural causes) we expect the threshold for deep convection to vary roughly with the tropical mean SST (e.g., Sobel et al. 2002) and the SST threshold for genesis (to the extent that such a thing exists) to vary similarly; thus we expect relative SST to be more appropriate than absolute SST for capturing the influence of climate variability on the probability of genesis. Recent studies show a strong empirical relation between relative SST and PI (Vecchi and Soden 2007a; Swanson 2008; Vecchi et al. 2008). PI, on the other hand, has the advantage that it includes atmospheric information in addition to SST that may influence the probability of TCG. However, being the theoretical maximum tropical cyclone intensity, PI, was not defined with the purpose of characterizing TCG, and while it is plausible to use it as a predictor, there is no strong theoretical basis for doing so. PI also has the disadvantage that it is highly derived quantity whose computation, compared to relative SST, is much more complex and requires more data. From a practical point of view, we find that using PI in place of relative SST in the Poisson regression model gives a higher AIC value (line 3 of of Table 1)[1].

We examine the log-linear dependence assumption by adding powers and products of the climate variables to the regression. Adding the 10 possible quadratic powers and products of the 4 climate variables one at a time to the Poisson regression, we find that including the square of the absolute vorticity most reduces AIC (line 4 of Table 1). The resulting negative coefficient (-0.17) for the square of absolute vorticity (line 4 of Table 1) means that in this model, increases in absolute vorticity reduce the expected number of TCG events for sufficiently large values of absolute vorticity. This behavior can be understood as the regression attempting to accommodate the lack of TCG events at higher latitudes where values of absolute vorticity are high on average. However, while such a dependence may fit the data, it would not appear to have a physical basis; we do not have a physical reason for associating the reduction in TCG occurrence at higher latitudes with increased absolute vorticity. Rather, a physical explanation would be that the lack of TCG events at high latitudes is due to insufficient ocean thermal energy there. A more attractive explanation of the results is that $\log \mu$ has a nonlinear dependence on absolute vorticity, and including the square of the absolute vorticity in the regression approximates that dependence. Including additional powers of absolute vorticity, as in a series expansion, could give a more physically satisfying dependence. However, including more powers of absolute vorticity in the index would increase its complexity

[1]We note that ERA PI is systemically larger than NCEP PI by about 38%; to fit NCEP PI and ERA PI simultaneously in the Poisson regression model, we scale ERA PI by the factor 0.72.

and make the estimation of its parameters less robust.

To further examine the functional dependence of the number of TCG events on absolute vorticity, we fit the Poisson regression for different ranges of absolute vorticity values. Specifically, for a given value η' of the absolute vorticity, we fit the Poisson regression using data in the range $\eta' - \delta\eta \leq \eta \leq \eta' + \delta\eta$ and thus obtain regression coefficients that depend on η'. To the extent that the coefficient b_η depends on η', the sensitivity of TCG events to changes in absolute vorticity depends on the value of absolute vorticity, and $\log \mu$ has a nonlinear dependence on absolute vorticity. This procedure is essentially equivalent to computing the partial logarithmic derivative of the number of TCG events with respect to absolute vorticity. The dependence of b_η on absolute vorticity is shown in Fig. 2 computed using $\delta\eta = 0.5$; the coefficient error bars are based on the estimated dispersion. For modest values of the absolute vorticity, b_η has significant positive values indicating that TCG increases with increasing absolute vorticity. For larger values of the absolute vorticity, b_η is not significantly different from zero, indicating that for this range of values, TCG is insensitive to further increases in absolute vorticity. Specifically, for absolute vorticity greater than about 4, further increases in absolute vorticity do not increase the expected number of TCG events. The value of b_η obtained using all values of the absolute vorticity (dark dashed line) is roughly the average of the values of b_η conditioned on η. This observed dependence of b_η on the value of η motivates our decision to use the quantity $\min(\eta, 3.7)$ rather than absolute vorticity in the index; we refer to this quantity as the clipped absolute vorticity. Generalized additive models provide a systematic method for including more complex functional dependence (Mestre and Hallegatte 2009; Villarini et al. 2010). The threshold value 3.7 was chosen to maximize the likelihood of the observations and therefore counts in the AIC as a parameter. Although the number of parameters is increased, AIC is smaller (line 5 of Table 1) for the Poisson regression based on the quantity

$\min(\eta, 3.7)$. The coefficient of $\min(\eta, 3.7)$ is 1.22 which is larger than that (0.58; line 2 of Table 1) of η. We will see later that this feature means that the index based on clipped absolute vorticity responds more strongly to the equator-ward decrease in absolute vorticity without the undesired side-effect of generating too many TCG events at high latitudes.

Adding additional powers and products to this set of climate variables does not reduce AIC. So, we take this model (line 5 of Table 1) as our TCG index and explore its properties in more detail in the next section.

4. Properties of the index

a. Climatological spatial distribution and basin integrated values

We now examine the properties of the TCG index developed in the previous section, focusing on physically relevant characteristics such as spatial distributions and basin-integrated values. Many of the important features of the index can be seen by examining its annually integrated values shown in Figure 3. For the most part, there is reasonable agreement between the observations and the indices both in spatial structure and in magnitude. However, the magnitude of the NCEP TCG index is weak in the Atlantic main development region. The region of maximal observed density of TCG events in the Northern Pacific extends further equatorward than is seen in either of the indices. Neither of the indices match the observed values in the Arabian Sea and the southern part of the North Indian ocean. The values of the indices in the South Pacific are too large compared to observations.

We examine the seasonal cycle of the TCG index by comparing the basin-integrated climatology of the TCG index with that of the observations; basins domains are defined in Table 2. Figure 4 shows that the TCG index captures the overall TCG seasonal structure in all basins to some extent. No scaling is applied to the basin-integrated TCG index values; the differing basin sizes and differing areas of grid points at different latitudes are included as an offset in the Poisson regression as described previously. For the most part, TCG indices based on the NCEP and ERA data have similar properties. In the southern hemisphere, the TCG index representation of the active season is not active enough, and its representation of the inactive season is too active. Failure to capture the peak activity is seen in the South Indian and Australian basins; peak activity is overestimated in the South Pacific basin consistent with Fig. 3. TCG index values are too large in all of the Southern basins during austral winter. The two peaks in the North Indian seasonal cycle are reproduced but with insufficient magnitude. In the Central North Pacific, the TCG index overestimates the amplitude during active season and has its maximum value about a month too late in the calendar year. The peak amplitude of the TCG index in the Eastern North Pacific is too weak and phased about a month too late in the calendar year.

Figure 4 also shows the hemispheric and basin integrated number of observed storms and index when the reanalysis relative humidity is used. In this case, the ERA-based index is too active during the peak seasons in the Northern hemisphere, especially during August in the Western North Pacific, Central North Pacific, and Atlantic. In the Southern hemisphere totals, the ERA relative humidity-based index is comparable to observed while NCEP relative humidity-based index is too weak. This behavior can be understood from the differing seasonal cycles of NCEP and ERA relative humidity shown in Fig. 1. Interestingly, the ERA relative humidity-based index does better reproduce the pre-monsoon TCG peak observed in the North Indian basin.

The observed and modeled number of TCG events per year for the January-March (JFM) and August-October (ASO) seasons using NCEP and ERA data, respectively, are shown in Figs. 5a,b and 6a,b. Peak season spatial distributions of the TCG index are similar to observations. ERA has a more active development region in the Atlantic. NCEP has a more active Western North Pacific basin. The negative bias of the TCG index in the Eastern North Pacific and positive bias in the South Pacific is apparent in both data sets. Figures 5e,f and 6e,f show the impact of using reanalysis relative humidity. When using NCEP relative humidity, the NCEP based index is weaker in the Western and Eastern North Pacific regions. In contrast, the ERA-based index is stronger in most basins when ERA relative humidity is used.

Using PI in the Poisson regression rather than relative SST leads to basin-integrated index amplitudes that are too low in the Northern Hemisphere and phased too late in the Southern Hemisphere as shown in Fig. 7. This phasing problem is seen in all Southern basins. The problem of low basin-integrated amplitude is worst in the Western North Pacific peak season. Figures 5(g,h) and 6(g,h) show that the impact of using PI on the spatial pattern is mostly in its amplitude, with overall index amplitudes being too high in the Southern Hemisphere and too low in the Northern Hemisphere.

If absolute vorticity rather than clipped absolute vorticity is used in the index, the Northern Hemisphere integrated August value is too high, primarily due to its being too high in the Western North Pacific and Atlantic as shown in Fig. 8. Using absolute vorticity rather than clipped absolute vorticity reduces the overall Southern Hemisphere peak values due to reduction in the South Indian and Australian basins. Using absolute vorticity rather than clipped absolute vorticity also leads to a shift in the phasing of the South Pacific seasonal cycle. Figures 5(i,j) and 6(i,j) show that if absolute vorticity rather than clipped absolute vorticity is used, the TCG index is too large on the equator and extends too far northward in the Atlantic and Western North Pacific during ASO. Further spatial details of the ASO field are shown in Fig. 9. Use of the clipped absolute vorticity improves the spatial pattern in the Southern Hemisphere during JFM by shifting positive values of TCG off of the equator and narrowing the spatial distribution.

We now compare the new TCG index with the GPI from Camargo et al. (2007a). As mentioned before the NCEP PI and ERA PI have systemically different amplitudes and therefore, so do the GPI values which depend on the third power of PI. We find separate multiplicative constants for the NCEP GPI and ERA GPI so that the area-weighted GPI best fits the observed number of TCG events. Figure 10 shows that overall GPI is too low during the Southern Hemisphere active season and too high during the inactive season. Northern Hemisphere integrated values of GPI are better, but looking at individual basins reveals that the Atlantic is too high during the active season while the Western North Pacific is too low. The GPI peak occurs during the wrong month in the South Indian basin. Australian GPI values are too low during the active season and too high during the inactive season. Figures 5(k,l) and 6(k,l) show that GPI spatial patterns extend too far poleward and are too close to the equator. Amplitude variability of the GPI is less than that of the new TCG index.

b. Dependence on climate variables

A method for examining the dependence of observed TCG and the TCG index on the individual climate variables is to compute "marginal" functions of a single variable. Marginal functions are constructed by averaging over all the variables except one. For instance, we define the marginal function $N_\eta(\eta')$ for absolute vorticity by

$$N_\eta(\eta') = \langle N(\eta_i, \mathcal{H}_i, T, V_i) \rangle \ , \qquad \text{where } i \text{ satisfies: } \eta_i - \delta \le \eta' \le \eta' + \delta \, ; \tag{10}$$

$\langle \cdot \rangle$ denotes average. Analogous marginal functions can be defined for relative humidity, relative SST and vertical shear. In the Poisson regression model the number of TCG events has a log-linear dependence on the climate variables. However, the dependence of the marginal function on the individual variables may not be log-linear due to correlations between the climate variables. Most of the correlations and hence much of the behavior of the marginal functions can be explained from the latitude dependence of the zonally averaged climate variables.

Figure 11a shows the dependence of $N_\eta(\eta')$ on vorticity as well as a histogram of the values of vorticity. In all the marginal function calculations, the range of the variable in question is divided into 50 equally spaced bins. For small values of absolute vorticity, the marginal function is an increasing function of absolute vorticity. For vorticity greater than about 4, it is a decreasing function. The explanation for this latter behavior is primarily the fact that absolute vorticity increases as one moves poleward, and therefore, on average, the largest absolute vorticity values are found at high latitudes where low SST values make TC formation unlikely. The dashed line shows the behavior for a model based on absolute vorticity rather than clipped absolute vorticity. Not using clipped absolute vorticity leads to to an index which does not respond strongly enough to small values of absolute vorticity near the equator and responds too strongly to large values of absolute vorticity.

The marginal function for relative humidity is mostly an increasing function of relative humidity (Fig. 11b). However, the number of TCG events decreases for relative humidity near 80% and there are no TCG events for higher values of relative humidity. This behavior can be understood by noting that relative humidity has its largest values on average near the equator and at high latitudes, both regions where there are few TCG events. What seems to be a nonlinear (in log) dependence is really a reflection of the correlation of relative humidity with other variables; near the equator relative humidity is on average a decreasing function of latitude while absolute vorticity is increasing. At high latitudes, relative humidity is an increasing function of latitude while relative SST is decreasing.

The marginal function for relative SST is an increasing function of relative SST except for the very highest values of relative SST where there are no TCG events (Fig. 11c). This regime corresponds to locations near the equator where absolute vorticity is small and there are few TCG events.

For very small values of vertical shear, the marginal function for vertical shear is almost flat as vertical shear increases (Fig. 11d). This behavior is explained by the fact that in the tropics zonally averaged vertical shear is an increasing function of latitude, thus reductions in vertical shear may correspond to moving closer to the equator and not lead to more TCG events. As vertical shear increases further, the marginal function decreases. The "heavy tail" poorly described by the Poisson model may be due to subtropical storms that can form in a high shear environment (Evans and Guishard 2009; Guishard et al. 2009) with many having baroclinically instabilities precursors that lead to TCG (Davis and Bosart 2003).

c. Decomposition of the seasonal cycle

The form of the index as the exponential of a sum of factors allows easy quantification of the importance of each of the factors. Figure 12 shows the seasonal cycle of the basin-averaged individual factor anomalies. In an overall sense, relative SST variation has the biggest impact on the seasonal cycle followed by vertical shear and finally relative humidity; clipped absolute vorticity has little contribution to the seasonal cycle. In the Eastern North Pacific, vertical shear has a larger contribution to the seasonal cycle than does relative SST. For the most part the seasonal cycle of the three factors is in phase. However, the behavior in the North Indian basin with its two maxima is more complicated. There, the reduction in vertical shear results in the pre-monsoon maximum. The increase in relative humidity following the start of the monsoon is offset by increases in vertical shear. The reduction in vertical shear and continued increased relative humidity results in the second maximum. It should noted that the TCG index in the North Indian basin, while capturing the two peaks, does not reproduce the observed magnitudes.

d. Interannual variability

We now examine how well the TCG index developed with climatological data can reproduce basin averaged interannual variability. We consider the period 1982-2001, a period when TCG observations are good. Since SSM/I data is not available during this period, we use an index based on reanalysis relative humidity. The Poisson regression model is fit to climatological data (line 6 Table 1) and applied to interannual data.

Tables 3 and 4 show the correlation between the observed and modeled basin-integrated seasonal (3-month) total number of TCG events based on NCEP and ERA data, respectively. Correlations were computed only for seasons whose average number of TCG is greater than one; correlation is a poor measure of association for Poisson variables with small expected values. Monte Carlo experiments show that the 95% significance level for correlation between two Poisson variables depends strongly on their mean value when the mean value is less than one. As the mean value increases, the Poisson variables become approximately Gaussian and the 95% significance level for correlation approaches that for Gaussian distributed variables which for sample size 20 is

0.377. Here, we conservatively consider correlations greater than 0.4 to be significant. The NCEP-based index has 23 season-basins with significant correlations while the ERA-based index has 25. For the most part, the TCG indices based on the two data set show similar correlation levels and seasonality. Neither the NCEP or the ERA index show any significant interannual correlation in the North Indian basin. However, NCEP has some significant correlations in the Australian region during austral summer while ERA does not. Also, ERA shows some significant correlation during the early part of the South Indian active period while NCEP does not. The NCEP index shows significant correlations in the western North Pacific during the start and end of the peak season while the ERA index shows significant correlations throughout the boreal summer. Correlation levels are roughly comparable to, though generally higher than, those found in Camargo et al. (2007a) using GPI; the 23 (25) significant correlations between the NCEP (ERA) based index and observations exceed the GPI ones in all but 3 (5) cases.

Much of the interannual variability is related to ENSO. The common period for the NCEP reanalysis and ERA is 1958-2002. For the purpose of compositing, the 11 years with the highest values of the three-month NINO3.4 index were classified at El Niño years and the 11 years with the lowest values as La Niña years. Table 5 shows the years selected for the composites. Figure 13 shows the NCEP and ERA based El Niño - La Niña composite difference maps for JFM and ASO. In JFM the El Niño - La Niña composite shows a decrease in the Western North Pacific and an increase in the Central North Pacific, an equator-ward shift in the South Pacific, and a decrease in the Australian basin. In ASO the El Niño - La Niña composite shows a reduction in the North Indian, Western North Pacific and Atlantic basins, and increases in the Central North Pacific and Eastern Pacific basins. As shown in Camargo et al. (2007a) for GPI, these shifts in the TCG ENSO composites reflect well the observed TCG behavior in the various basins for JFM and ASO.

5. Discussion: the role of vorticity

The superior performance of the clipped vorticity relative to vorticity itself is an unexpected result which may have some significance for our understanding of the physics of the genesis process. It is well known that a finite background low-level absolute vorticity is necessary to the genesis process, as is immediately evident from the fact that genesis almost never occurs within a few degrees of the equator. It is less clear what one ought to expect the dependence of the probability of genesis on the value of the background vorticity to be. It is plausible, but by no means obvious, that once the vorticity reaches some sufficient value, it no longer is a rate-limiting factor and other aspects of the environment (such as thermodynamic parameters and vertical shear) become more critical. Our results suggest that this is in fact the case.

This conclusion is most likely dependent on the choice of averaging time used to define the environmental fields. On daily time scales, it is certain that the existence of a pre-existing tropical depression makes genesis more likely compared to the absence of a depression, and it seems likely that the vorticity of the depression is one of the factors that makes it so. Nonetheless, inasmuch as it is useful to quantify the probability of genesis based on monthly mean fields, our finding indicates that the probability of genesis does not increase further with low-level absolute vorticity once that variable reaches a threshold value.

6. Summary and conclusions

The likelihood of tropical cyclone genesis (TCG) is observed to depend on features of the large-scale climate. Therefore changes in climate due to either natural or anthropogenic causes can lead to changes in the likelihood of TCG. Given the incompleteness of the theoretical understanding of TCG, empirical indices are a useful way of encapsulating observed relations between TCG and large-scale climate variables. Here, in the spirit of earlier work (Gray 1979; DeMaria et al. 2001; Royer et al. 1998; Emanuel and Nolan 2004), we construct a TCG index which is a function of climate variables and whose size reflects the probability of genesis.

We construct the index by developing a Poisson regression between the observed monthly number of storms over a 40-year period and the monthly climatological values of the climate variables. An attractive feature of this approach is that it is objective and hence easily applicable to other data sets. Moreover, the regression methodology provides a natural framework for selecting the variables to be used in the index and assessing the performance of the index. Initially we take as predictors in the index: absolute vorticity at 850 hPa, 600 hPa reanalysis relative humidity, relative SST and vertical shear between the 850 hPa and 200 hPa levels; relative SST is the difference between the local SST and the mean tropical SST

The Poisson regression assumes a log-linear relation between the number of TCG events and the climate variables. This assumption is equivalent to assuming that the sensitivity, as measured by the logarithmic partial derivative, of the number of storms to changes in the individual climate variable is constant. We find that the data do not support this assumption for absolute vorticity. In particular, the sensitivity of the number of TCG events to absolute vorticity is roughly constant and nonzero for values of absolute vorticity less than $4 \times 10^5 s^{-1}$, while for values of absolute vorticity greater than $4 \times 10^5 s^{-1}$, it is close to zero. This property of the data suggests the use of the “clipped” absolute vorticity, defined as the absolute vorticity itself when that quantity is below a threshold and the threshold value otherwise. We find that use of the clipped absolute vorticity in the TCG index improves the fit of the regression and results in more realistic spatial distributions with fewer TCG events near the equator and at high latitudes. Besides the practical value of this result for improving the performance of the index, it suggests a physical interpretation which is relevant to our understanding of the genesis process: while greater low-level ambient (monthly mean) vorticity increases the probability of genesis up to a point, beyond that point it does not continue to do so. While it is likely that increases in local vorticity on the daily time scale would still be a positive factor in genesis, our result may indicate that on a monthly mean basis, once vorticity is sufficiently large it tends to be the case that other factors (either thermodynamics or vertical shear) become rate-limiting.

We make a limited exploration of some alternative predictors in the index. We examine the impact of using SST or potential intensity (PI) rather than relative SST and of using satellite based column integrated relative humidity rather than the reanalysis products. Using climatological data, relative SST and SST contain nearly the same information, and indices based on them are very similar. However, relative SST allows better for changes in the mean climate, though perhaps still not optimally. PI also depends on the mean climate and has been observed to be well-correlated with relative SST (Vecchi and Soden 2007a; Swanson 2008), a result with a straightforward physical basis (Sobel et al. 2002; Ramsay and Sobel 2010). However, relative SST is considerably simpler to compute and understand than PI, and in fact, we find that relative SST performs better in the index than PI. The relative advantage of using satellite-based column-integrated is mixed compared to using reanalysis humidity at a single level, with a modest positive (negative) impact seen with respect to the NCEP (ERA) reanalysis product.

Overall the TCG index reproduces much of the observed basin-integrated seasonality and spatial patterns. The index also reproduces well the marginal dependence of the number of TCG events on the individual climate variables. This dependence is not log-linear due to correlations between variables. For the most part, the NCEP- and ERA-based indices have similar properties and common deficiencies. There are errors in the details of the spatial structure in the Northern Pacific. Index values are too small in the Arabian Sea. The observed number of TCG events in the South Pacific are smaller than that predicted by the index. The magnitude of the NCEP-based TCG index is weak in the Atlantic main development region. Comparison with the GPI, a TCG index based on PI (Emanuel and Nolan 2004; Camargo et al. 2007a) shows that the index developed here has better performance and avoids the complexity of associated with PI.

Developing the TCG index using climatological data results in an index that fits the climate-TCG co-variability contained in the seasonal cycle and in different geographical regions. Applying the index to interannually varying climate data, we show that the index is also able to reproduce some interannual variability and spatial shifts due to ENSO. Future work will apply the index to climate change scenarios.

Acknowledgments.

We thank Larissa Back and Chris Bretherton for making available the scripts for calculating the column-relative humidity. We especially thank Larissa Back for her help and explanations. We also thank Kerry Emanuel and Gabriel Vecchi for discussions. MKT is supported by a grant/cooperative agreement from the National Oceanic and Atmospheric Administration (NA05OAR4311004). SJC and AHS acknowledge support from NOAA Grant NA08OAR4320912. The views expressed herein are those of the authors and do not necessarily reflect the views of NOAA or any of its sub-agencies. The ECMWF ERA-40 data used in this study were obtained from the ECMWF data server.

REFERENCES

Akaike, H., 1973: Information theory and an extension of the maximum likelihood principle. *2nd International Symposiium on Information Theory*, B. N. Petrov and F. Czáki, Eds., Akademiai Kiadó, Budapest, 267–281.

Back, L. E. and C. S. Bretherton, 2009: A simple model of climatological rainfall and vertical motion patterns over the tropical oceans. *J. Climate*, **22**, 6477–6497.

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

Bister, M. and K. A. Emanuel, 1998: Dissipative heating and hurricane intensity. *Meteor. Atm. Phys.*, **52**, 233–240.

Bister, M. and K. A. Emanuel, 2002a: Low frequency variability of tropical cyclone potential intensity. 1. Interannual to interdecadal variability. *J. Geophys. Res.*, **107**, 4801, doi:10.1029/2001JD000776.

Bister, M. and K. A. Emanuel, 2002b: Low frequency variability of tropical cyclone potential intensity. 2. Climatology for 1982-1995. *J. Geophys. Res.*, **107**, 4621, doi:10.1029/2001JD000780.

Bretherton, C. S., M. E. Peters, and L. E. Back, 2004: Relationships between water vapor path and precipitation over the tropical oceans. *J. Climate*, **17**, 1517–1528.

Bye, J. and K. Keay, 2008: A new hurricane index for the Caribbean. *Interscience*, **33**, 556–560.

Camargo, S. J., K. A. Emanuel, and A. H. Sobel, 2007a: Use of genesis potential index to diagonose ENSO effects on Tropical cyclone genesis. *J. Climate*, **20**, 4819–4834.

Camargo, S. J., A. H. Sobel, A. G. Barnston, and K. A. Emanuel, 2007b: Tropical cyclone genesis potential index in climate models. *Tellus*, **59A**, 428–443.

Camargo, S. J., M. C. Wheeler, and A. H. Sobel, 2009: Diagnosis of the MJO modulation of tropical cyclogenesis using an empirical index. *J. Atmos. Sci.*, **66**, 3061–3074.

Cheung, K. K. W., 2004: Large-cale environmental parameters associated with tropical cyclone formations in the western north Pacific. *J. Climate*, **17**, 466–484.

Davis, C. A. and L. F. Bosart, 2003: Baroclinically induced tropical cyclogenesis. *Mon. Wea. Rev.*, **131**, 2730–2747.

DeMaria, M., J. A. Knaff, and B. H. Conell, 2001: A tropical cyclone genesis parameter for the tropical Atlantic. *Wea. Forecasting*, **16**, 219–233.

Elsner, J. B. and C. P. Schmertmann, 1993: Improving extended-range seasonal forecasts of intense Atlantic hurricane activity. *Wea. Forecasting*, **8**, 345–351.

Emanuel, K. A., 1989: The finite-amplitude nature of tropical cyclogenesis. *J. Atmos. Sci.*, **46**, 3431–3456.

Emanuel, K. A., 1995: Sensitivity of tropical cyclones to surface exchange coefficients and a revised steady-state model incorporating eye dynamics. *J. Atmos. Sci.*, **52**, 3969–3976.

Emanuel, K. A. and D. S. Nolan, 2004: Tropical cyclone activity and global climate. *Proc. of 26th*

Conference on Hurricanes and Tropical Meteorology, Miami, FL, American Meteorological Society, 240–241.

Evans, J. L. and M. P. Guishard, 2009: Atlantic subtropical storms. Part I: Diagnostic criteria and composite analysis. *Mon. Wea. Rev.*, **137**, 2065–2080.

Gray, W. M., 1979: *Meteorology over the tropical oceans*, chap. Hurricanes: Their formation, structure and likely role in the tropical circulation, 155–218. Roy. Meteor. Soc.

Gualdi, S., E. Scoccimarro, 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.

Guishard, M. P., J. L. Evans, and R. E. Hart, 2009: Atlantic subtropical storms. Part II: Climatology. *J. Climate*, **22**, 3574–3594.

Kalnay, E. and Coauthors, 1996: The NCEP/NCAR 40-Year Re-anlysis Project. *Bull. Am. Meteor. Soc.*, **77**, 437–471.

Kistler, R., et al., 2001: The NCEP-NCAR 50-Year Reanalysis: Monthly Means CD-ROM and Documentation. *Bull. Am. Meteor. Soc.*, **82**, 247–267.

Kotal, S. D., P. K. Kundu, and S. K. R. Bhowmik, 2009: Analysis of cyclogenesis parameter for developing and nondeveloping low-pressure systems over the Indian Sea. *Nat. Hazards*, **50**, 389–402.

Lyon, B. and S. J. Camargo, 2009: The seasonally-varying influence of ENSO on rainfall and tropical cyclone activity in the Philippines. *Clim. Dyn*, **32**, 125 – 141, doi:10.1007/s00382-008-0380-z.

McDonnell, K. A. and N. J. Holbrook, 2004: A Poisson regression model of tropical cyclogenesis for the Australian-Southwest Pacific Ocean region. *Wea. Forecasting*, **19**, 440–455.

Mestre, O. and S. Hallegatte, 2009: Predictors of tropical cyclone numbers and extreme hurricane intenisties over the North Atlantic using generalized additive and linear models. *J. Climate*, **22**, 633–648.

Murakami, H. and B. Wang, 2010: Future change of North Atlantic tropical cyclone tracks: Projection by a 20-km-mesh global atmospheric model. *J. Climate, in press*, doi:10.1175/2010JCLI3338.1.

Nolan, D. S., E. D. Rappin, and K. A. Emanuel, 2007: Tropical cyclogenesis sensitivity to environmental parameters in radiative-convective equilibrium. *Q. J. R. Meteorol. Soc*, **133**, 2085–2107.

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

Ramsay, H. A. and A. H. Sobel, 2010: The effects of relative and absolute sea surface temperature on tropical cyclone potential intensity using a single column model. *submitted, J. Climate*.

Royer, J.-F., F. Chauvin, B. Timbal, P. Araspin, and D. Grimal, 1998: A GCM study of the impact of greenhouse gas increase on the frequency of occurrence of tropical cyclone. *Climatic Change*, **38**, 307–343.

Sall, S. M., H. Sauvageot, A. T. Gaye, A. Viltard, and P. de Felice, 2006: A cyclogenesis index for tropical Atlantic off the African coasts. *Atmospheric Research*, **79**, 123–147.

Smith, T. M. and R. W. Reynolds, 2004: Improved extended reconstruction of SST (1854-1997). *J. Climate*, **17**, 2466–2477.

Sobel, A. H., I. M. Held, and C. S. Bretherton, 2002: The ENSO signal in tropical tropospheric temperature. *J. Climate*, **15**, 2702–2706.

Solow, A. and N. Nicholls, 1990: The relationship between the Southern Oscillation and tropical cyclone frequency in the Australian region. *J. Climate*, **3**, 1097–1101.

Swanson, K. L., 2008: Nonlocality of atlantic tropical cyclone intensity. *Geochem. Geophys. Geosyst.*, **9**, Q04V01.

Uppala, S. M., et al., 2005: The ERA-40 re-analysis. *Quart. J. Roy. Meteor. Soc.*, **131**, 2961–3012, doi:10.1256/qj.04.176.

Vecchi, G. A. and B. J. Soden, 2007a: Effect of remote sea surface temperature change on tropical cyclone potential intensity. *Nature*, **450**, 1066–1070.

Vecchi, G. A. and B. J. Soden, 2007b: Increased tropical Atlantic wind shear in model projections of global warming. *Geophys. Res. Lett.*, **34**, L08 702, doi:10.1029/2006GL028905.

Vecchi, G. A., K. L. Swanson, and B. J. Soden, 2008: Whither hurricane activity? *Science*, **322**, 687, doi:10.1126/science.1164 396.

Villarini, G., G. A. Vecchi, and J. A. Smith, 2010: Modeling of the dependence of tropical storm counts in the North Atlantic basin on climate indices. *Mon. Wea. Rev.*

Wentz, F. J. and R. W. Spencer, 1998: SSM/I rain retrievals within a unified all-weather ocean algorithm. *J. Atmos. Sci.*, **55**, 1613–1627.

Yokoi, S. and Y. N. Takayuba, 2009: Multi-model projection of global warming impact on tropical cyclone genesis frequency over the western North Pacific. *J. Meteor. Soc. Japan*, **87**, 525–538.

Yokoi, S., Y. N. Takayuba, and J. C. L. Chan, 2009: Tropical cyclone genesis frequency over the western North Pacific simulated in medium-resolution coupled general circulation models. *Clim. Dyn.*, **33**, 665–683.

Zhao, M., I. M. Held, 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.

List of Tables

		vorticity			humidity		heat		shear	
	b	b_η	b_{η^2}	$b_{\min(\eta,3.7)}$	$b_{\mathcal{H}}$	$b_{\mathcal{H}_{\mathrm{SSMI}}}$	b_T	b_{PI}	b_V	AIC
1	-4.52	0.53	-	-	0.05	-	0.67	-	-0.15	28018.69
2	-10.05	0.58	-	-	-	0.11	0.62	-	-0.13	27662.02
3	-16.04	0.63	-	-	-	0.07	-	0.19	-0.13	29757.65
4	-12.04	1.93	-0.17	-	-	0.11	0.48	-	-0.13	26247.05
5	-11.18	-	-	1.22	-	0.10	0.52	-	-0.12	$\boxed{26029.55}$
6	-6.09	-	-	1.15	0.05	-	0.60	-	-0.14	26324.44

TABLE 1. Coefficients of the Poisson regression between number of TCG events per area per 40 years and climate variables. The coefficient subscripts indicate the quantities they multiply. Entries with “-” indicate variables not included in the model.

Northern Hemisphere (NH)	0 - 40°N
Southern Hemisphere (SH)	0 - 40°S
South Indian (SI)	30 - 100 °E
Australian (Aus)	100-180 °E
South Pacific (SP)	180 -110 °W
North Indian (NI)	40-100 °E
Western North Pacific (WNP)	100°E - 180°
Central North Pacific (CNP)	180° - 140°W
Eastern North Pacific (EMP)	140°W to American coast
Atlantic (Atl)	American coast to African coast

TABLE 2. Domain definitions used for basin integrations.

	SI	Aus	SP	NI	WNP	CNP	ENP	Atl
JFM	-0.16	$\boxed{0.51}$	$\boxed{0.46}$	-	0.20	-	-	-
FMA	0.18	$\boxed{0.48}$	$\boxed{0.53}$	-	$\boxed{0.58}$	-	-	-
MAM	- 0.04	0.34	-	-0.02	$\boxed{0.57}$	-	-	-
AMJ	0.06	0.06	-	-0.08	$\boxed{0.44}$	-	0.23	-
MJJ	-	-	-	-0.11	$\boxed{0.63}$	-	0.04	0.09
JJA	-	-	-	-	$\boxed{0.52}$	0.06	0.06	0.27
JAS	-	-	-	-	0.37	0.38	0.33	$\boxed{0.57}$
ASO	0.36	-	-	0.06	0.39	$\boxed{0.61}$	$\boxed{0.53}$	$\boxed{0.56}$
SON	0.31	-	-	-0.18	0.37	$\boxed{0.79}$	$\boxed{0.68}$	0.39
OND	0.11	0.21	-	-0.11	$\boxed{0.51}$	-	0.39	$\boxed{0.43}$
NDJ	0.22	$\boxed{0.44}$	$\boxed{0.63}$	0.03	$\boxed{0.53}$	-	-	-
DJF	-0.01	$\boxed{0.44}$	$\boxed{0.50}$	-	$\boxed{0.54}$	-	-	-

TABLE 3. Correlations between seasonal basin-integrated observed number of TCG events and TCG index based on NCEP data. Seasons averaging less than one TCG event per year are not included. Correlations greater than 0.4 are boxed.

	SI	Aus	SP	NI	WNP	CNP	ENP	Atl
JFM	-0.33	0.24	0.46	-	0.20	-	-	-
FMA	0.31	0.25	0.53	-	0.56	-	-	-
MAM	-0.06	0.16	-	0.17	0.52	-	-	-
AMJ	0.00	-0.08	-	0.21	0.37	-	0.02	-
MJJ	-	-	-	0.21	0.75	-	0.05	0.40
JJA	-	-	-	-	0.60	0.05	0.13	0.58
JAS	-	-	-	-	0.67	0.25	0.31	0.57
ASO	0.62	-	-	0.07	0.61	0.60	0.41	0.55
SON	0.45	-	-	-0.18	0.62	0.79	0.41	0.30
OND	0.17	0.21	-	-0.27	0.48	-	0.30	0.44
NDJ	0.03	0.34	0.68	-0.16	0.52	-	-	-
DJF	-0.29	0.16	0.42	-	0.51	-	-	-

TABLE 4. As in Fig. 3 but for ERA data.

JFM		ASO	
El Niño	La Niña	El Niño	La Niña
1958	1968	1963	1964
1966	1971	1965	1967
1969	1974	1969	1970
1970	1976	1972	1971
1973	1985	1976	1973
1977	1986	1982	1974
1983	1989	1986	1975
1987	1996	1987	1978
1992	1999	1991	1988
1995	2000	1994	1998
1998	2001	1997	1999

TABLE 5. Years used to form ENSO composites.

List of Figures

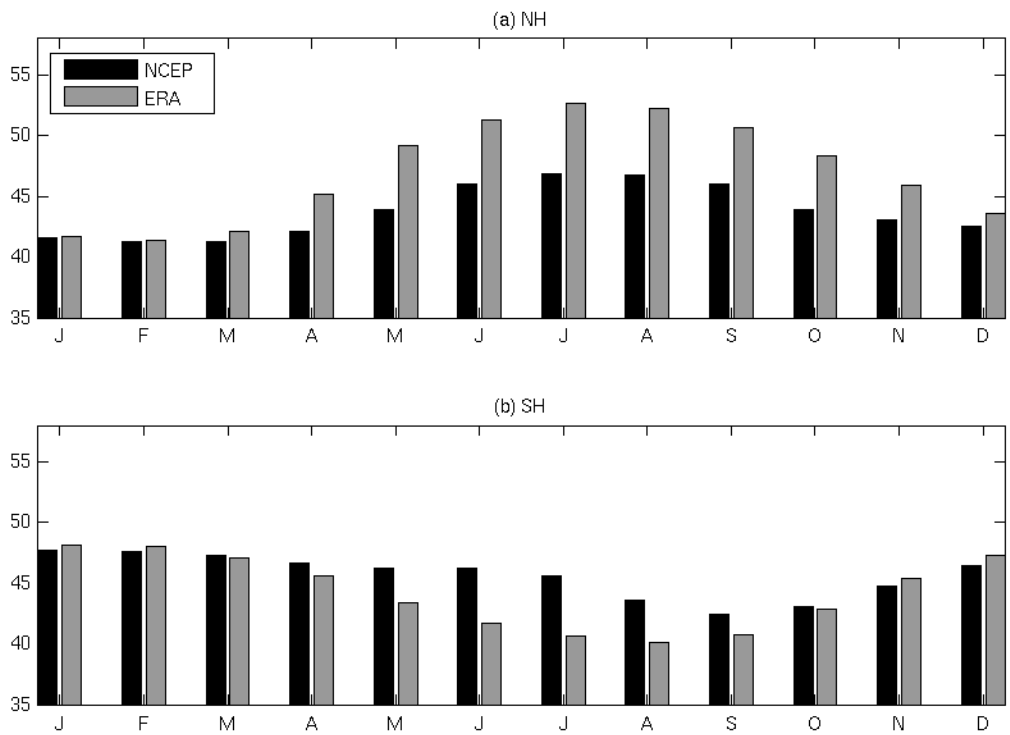

FIG. 1. Hemispheric (ocean grid points) averaged 600 hPa relative humidity from the NCEP reanalysis and ERA.

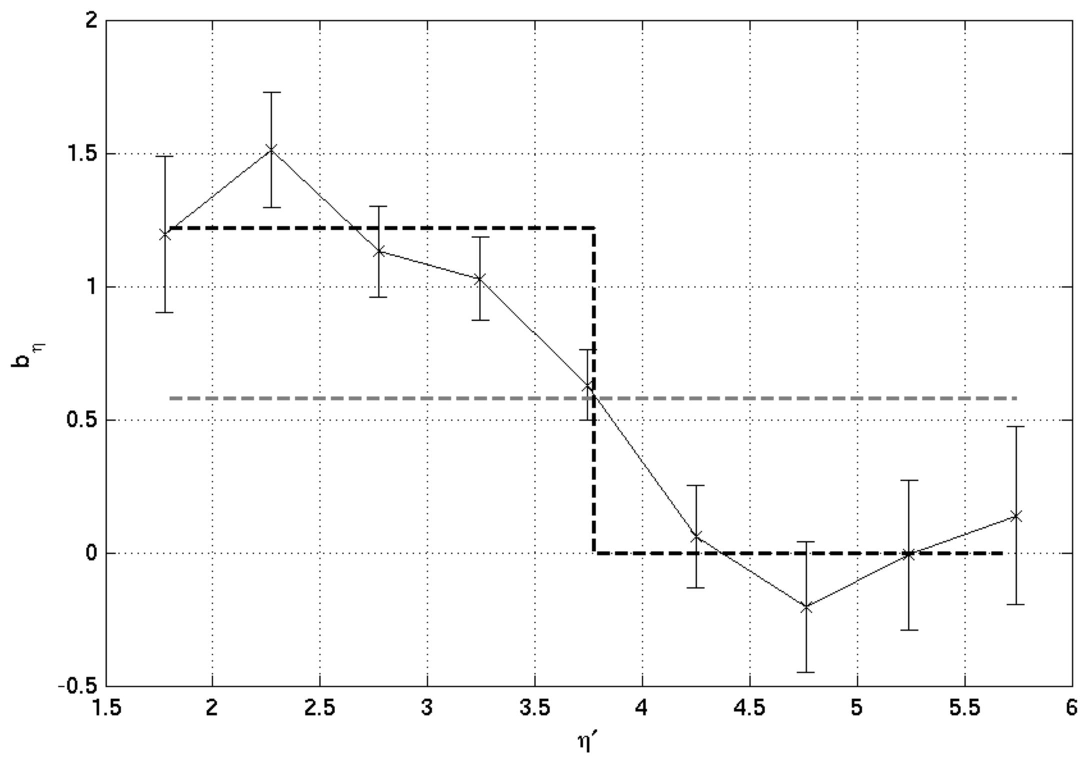

FIG. 2. The solid line shows b_η as a function of η' fit using values of the absolute vorticity in the interval $(\eta' - 0.5, \eta' + 0.5)$; error bars represent the 95% confidence intervals. The light dashed line is the value of b_η obtained using all the data. The dark dashed line corresponds to using $\min(\eta, 3.7)$ rather than η in the regression.

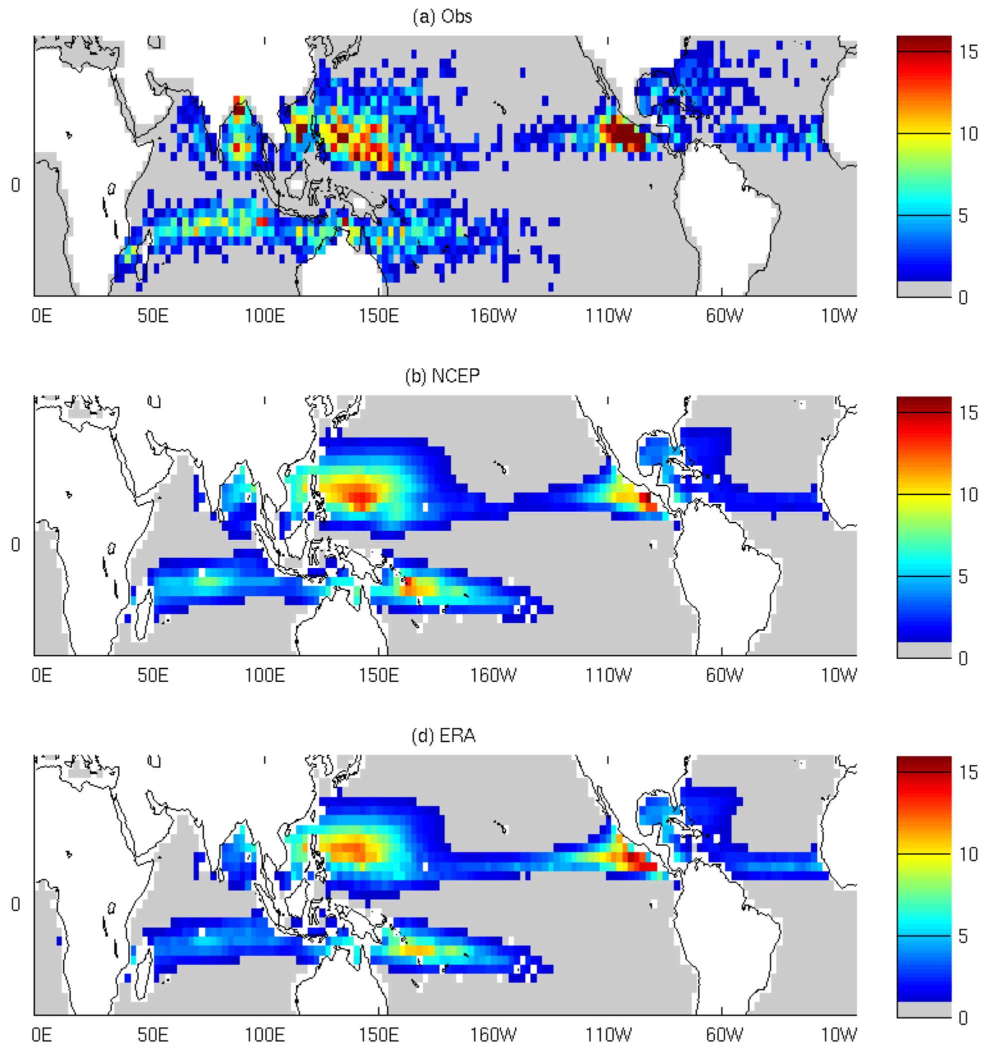

FIG. 3. Number of (a) observed TCG events over the 40-year climatology period and TCG indices based in (b) NCEP and (C) ERA data.

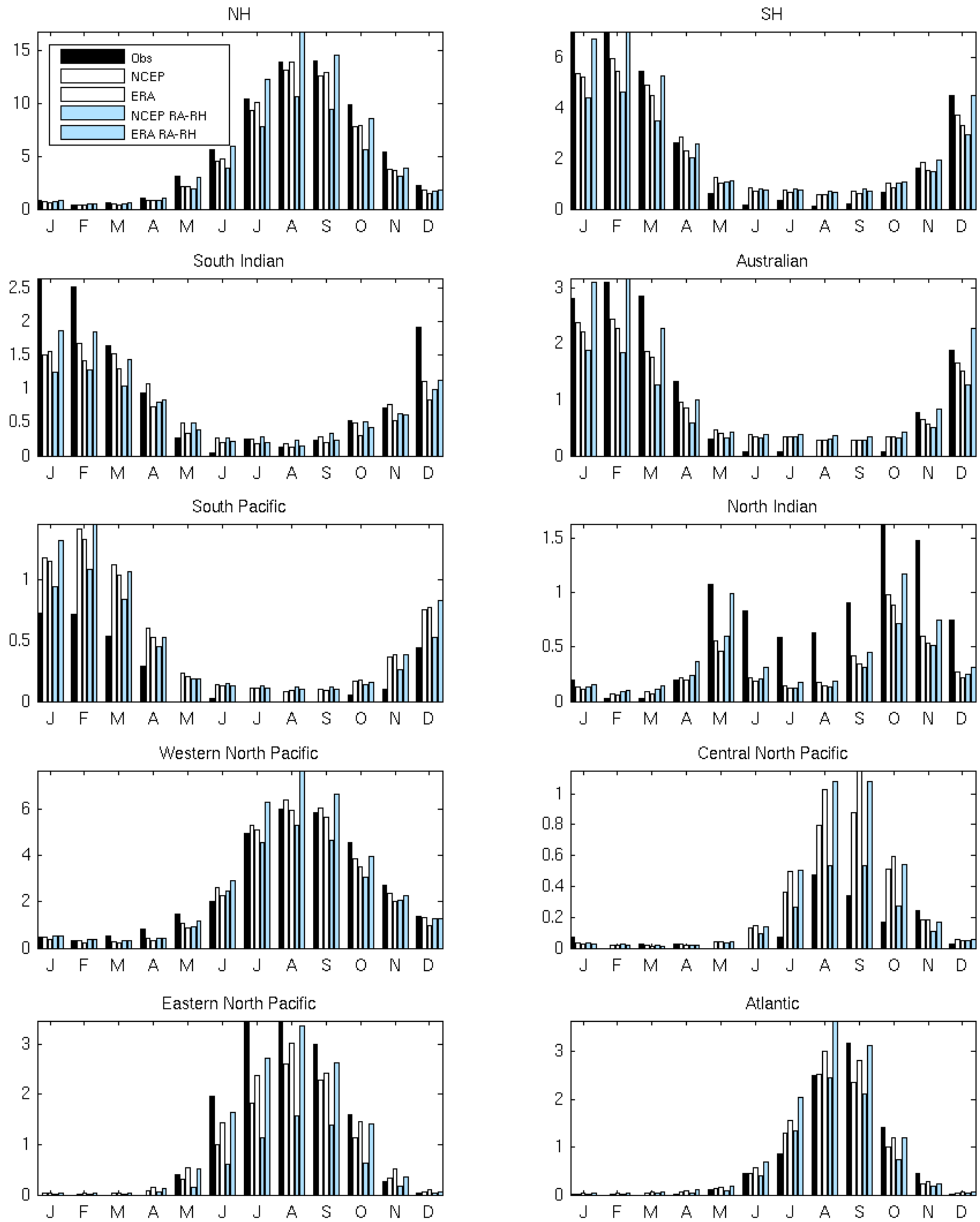

FIG. 4. Basin integrated climatology (number of TCG events per year). RA-RH indicates results from the Poisson regression model with reanalysis 600 hPa relative humidity rather than SSMI column-integrated relative humidity.

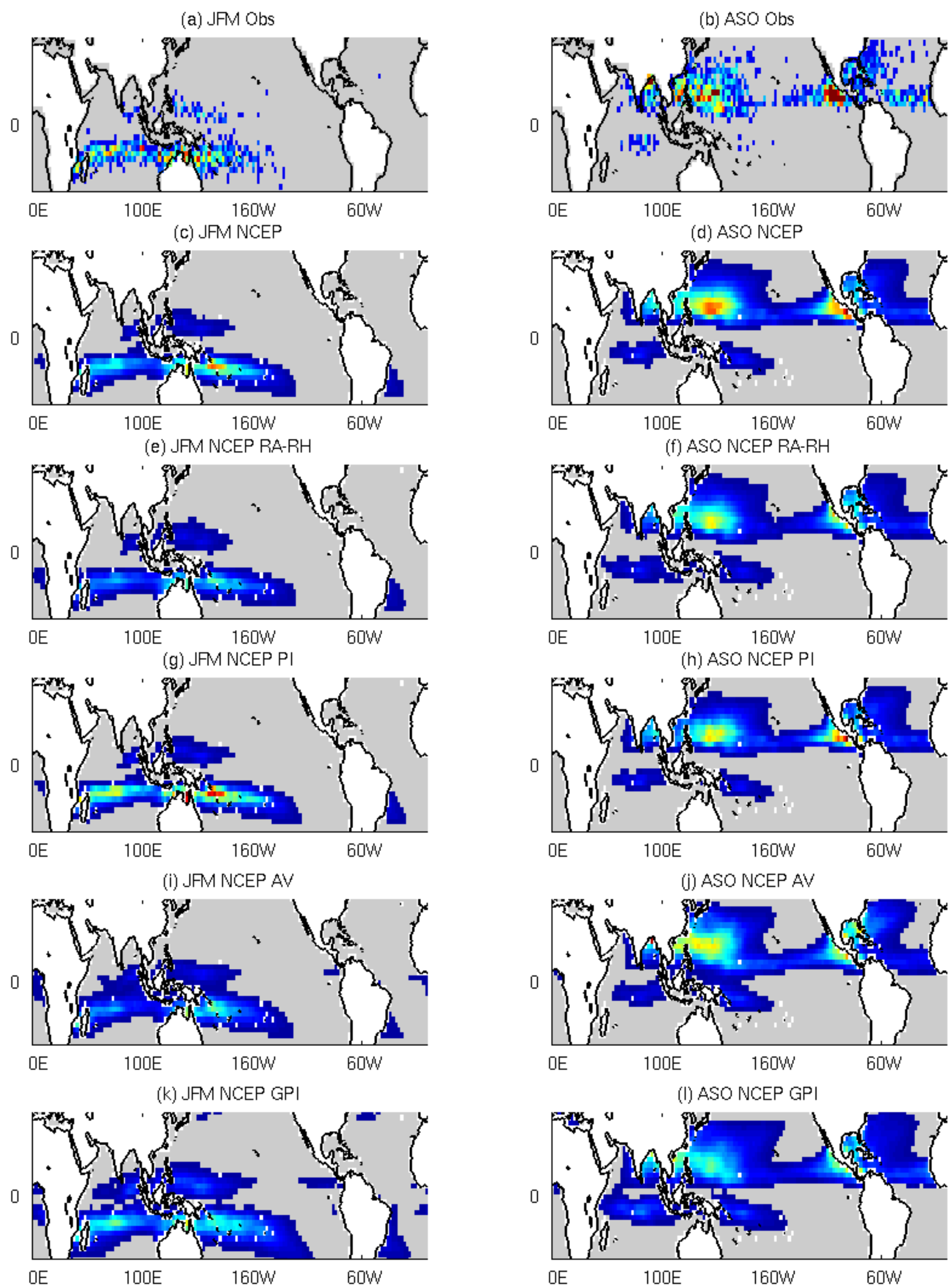

FIG. 5. Observed and modeled TCG peak season maps based on NCEP data. RA-RH indicates use of the reanalysis relative humidity; PI indicates use of PI rather than relative SST; AV indicates use of absolute vorticity rather than clipped absolute vorticity; GPI indicates use of GPI.

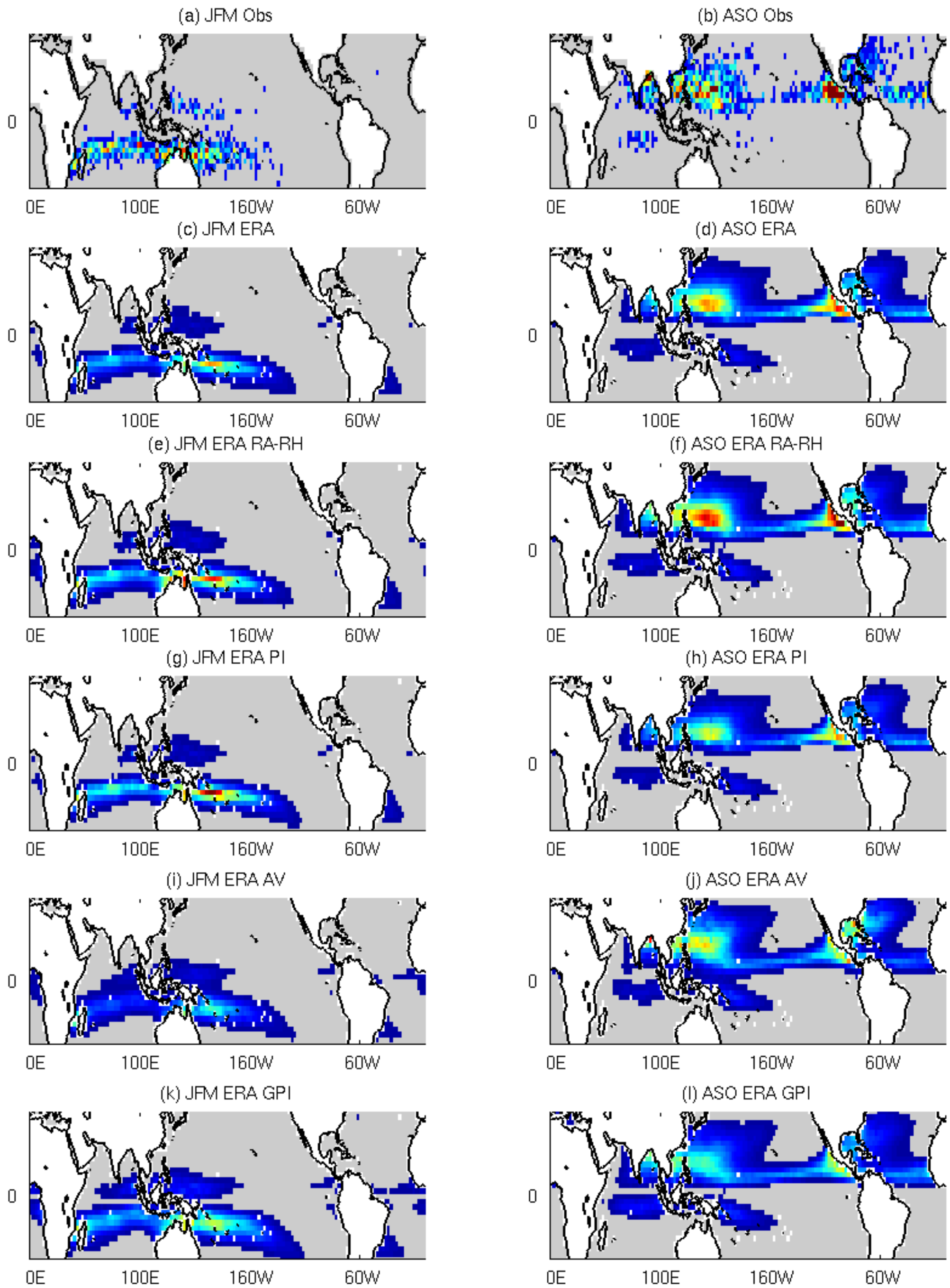

FIG. 6. As in Fig. 5 but for ERA data.

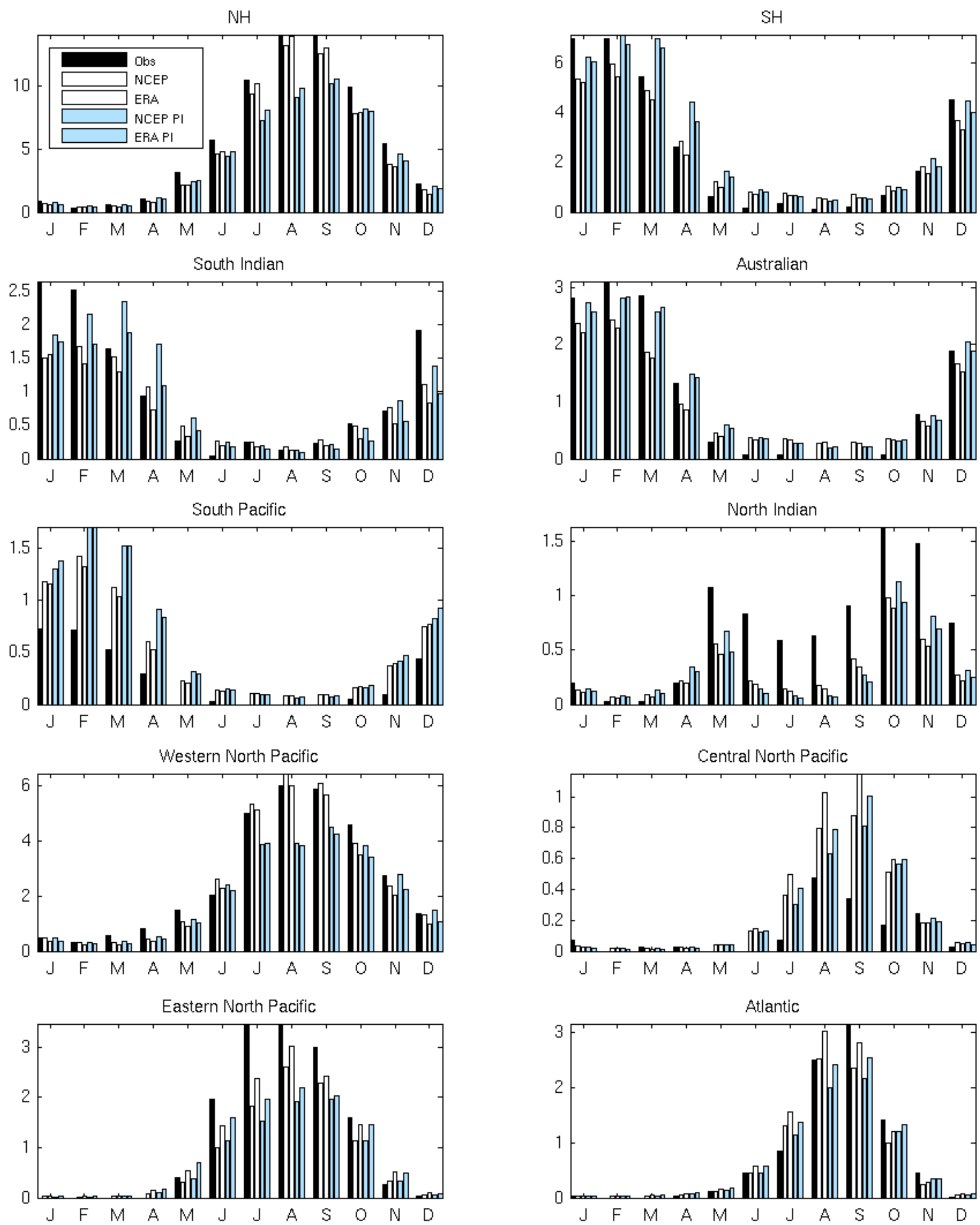

FIG. 7. Basin integrated climatology (number of TCG events per year). PI indicates results from the Poisson regression model with PI rather than relative SST.

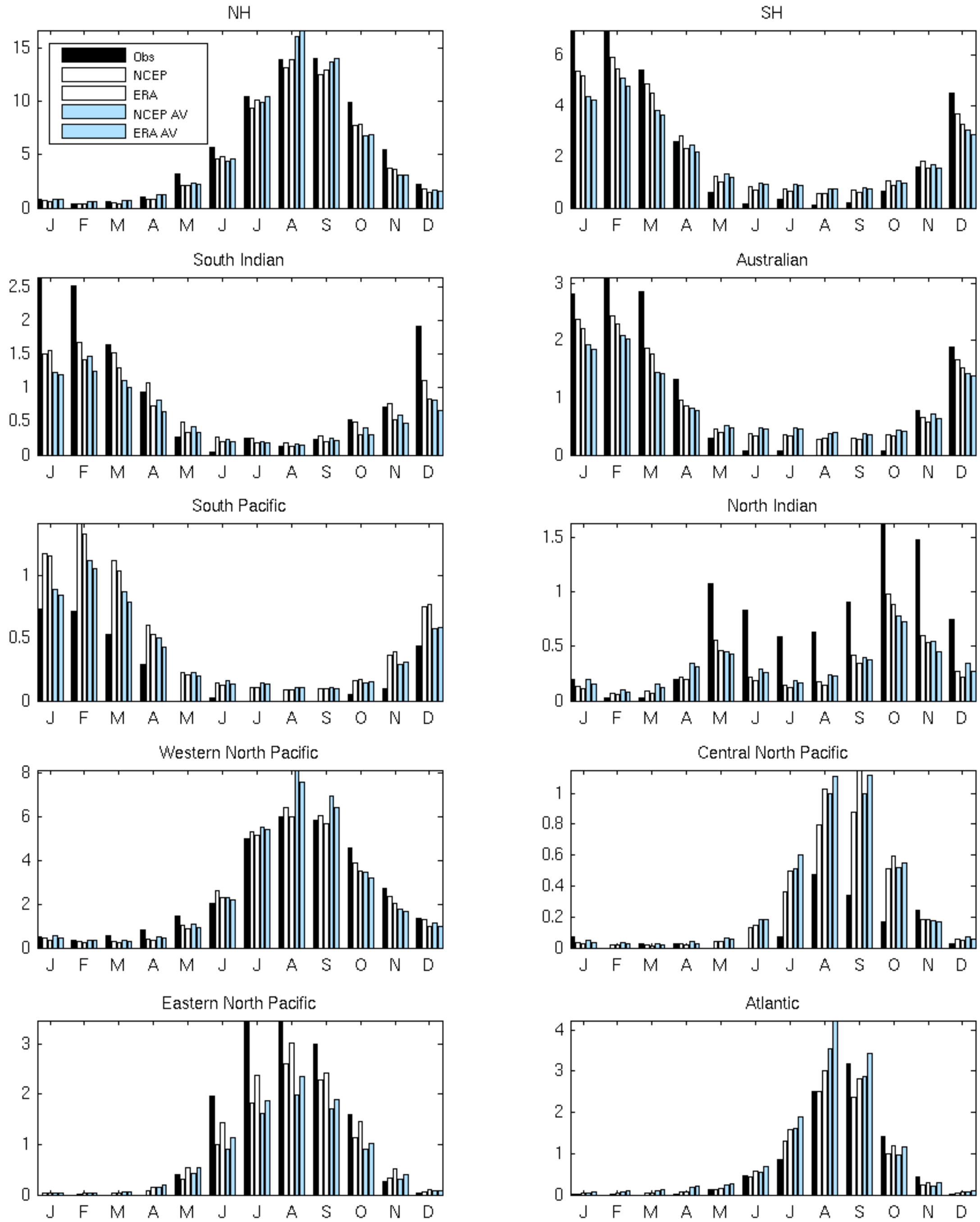

FIG. 8. Basin integrated climatology (number of TCG events per year). AV indicates results from the Poisson regression model with absolute vorticity rather than clipped absolute vorticity.

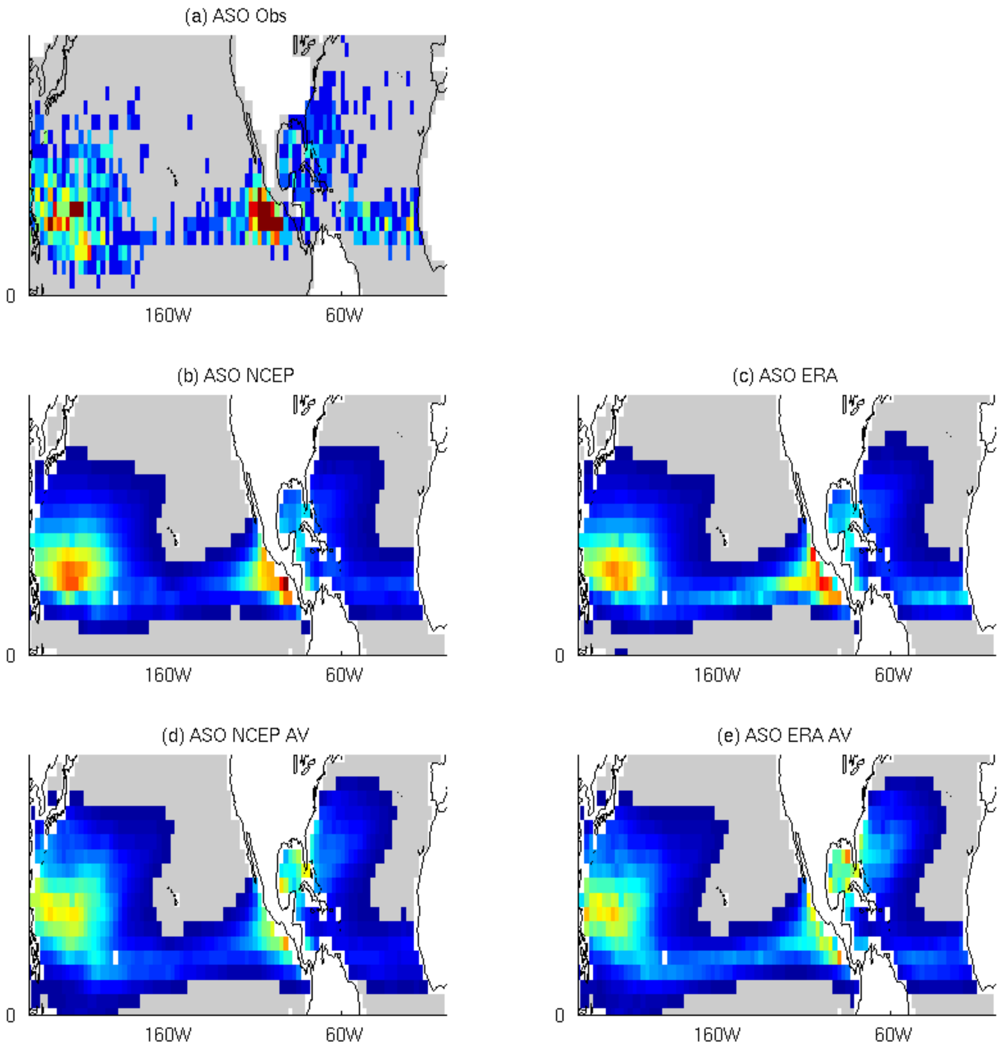

FIG. 9. Zoom in of Figures 5(i,j) and 6(i,j).AV indicates results from the Poisson regression model with absolute vorticity rather than clipped absolute vorticity.

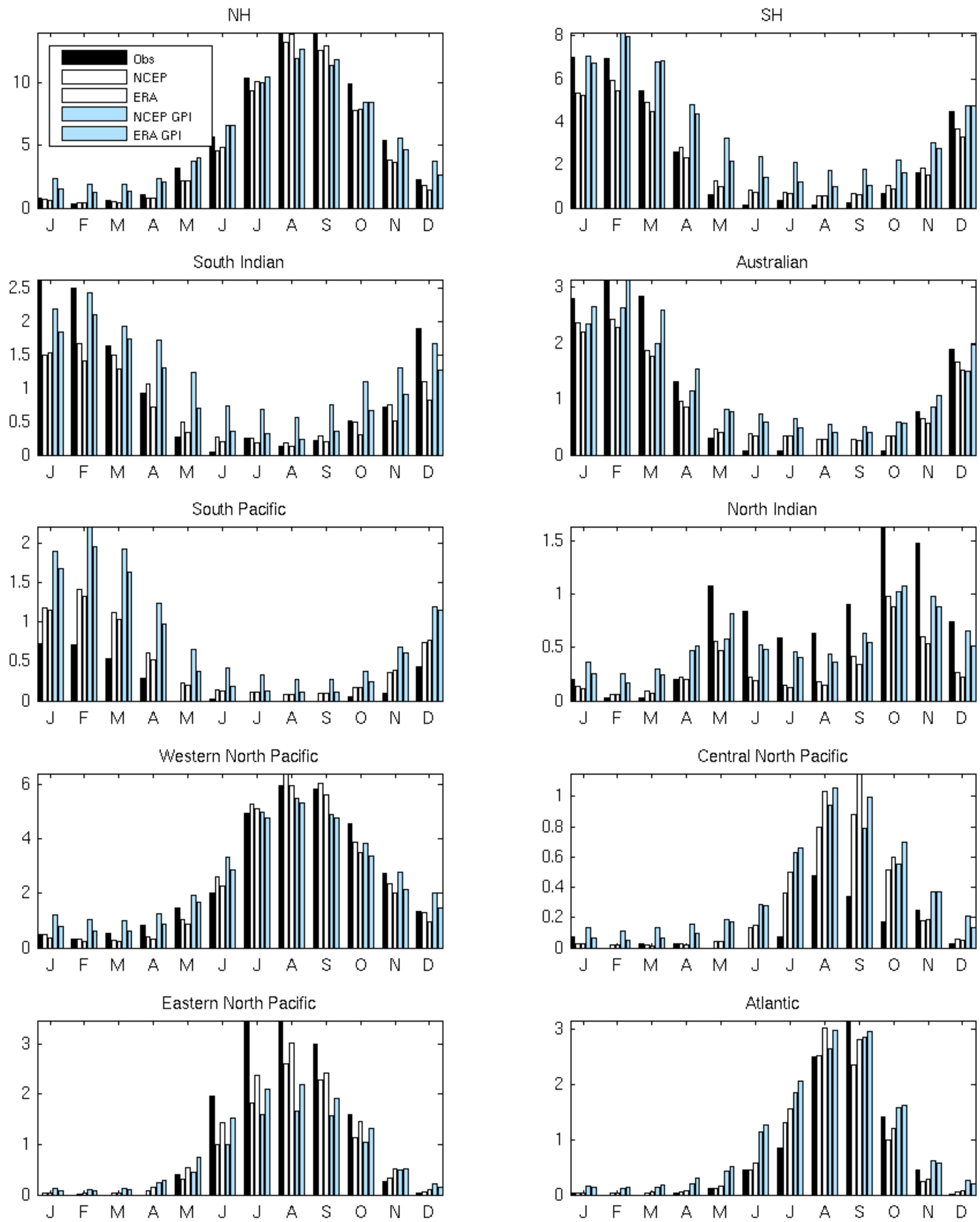

FIG. 10. Basin integrated climatology (number of TCG events per year). GPI indicates area-weighted and scaled GPI.

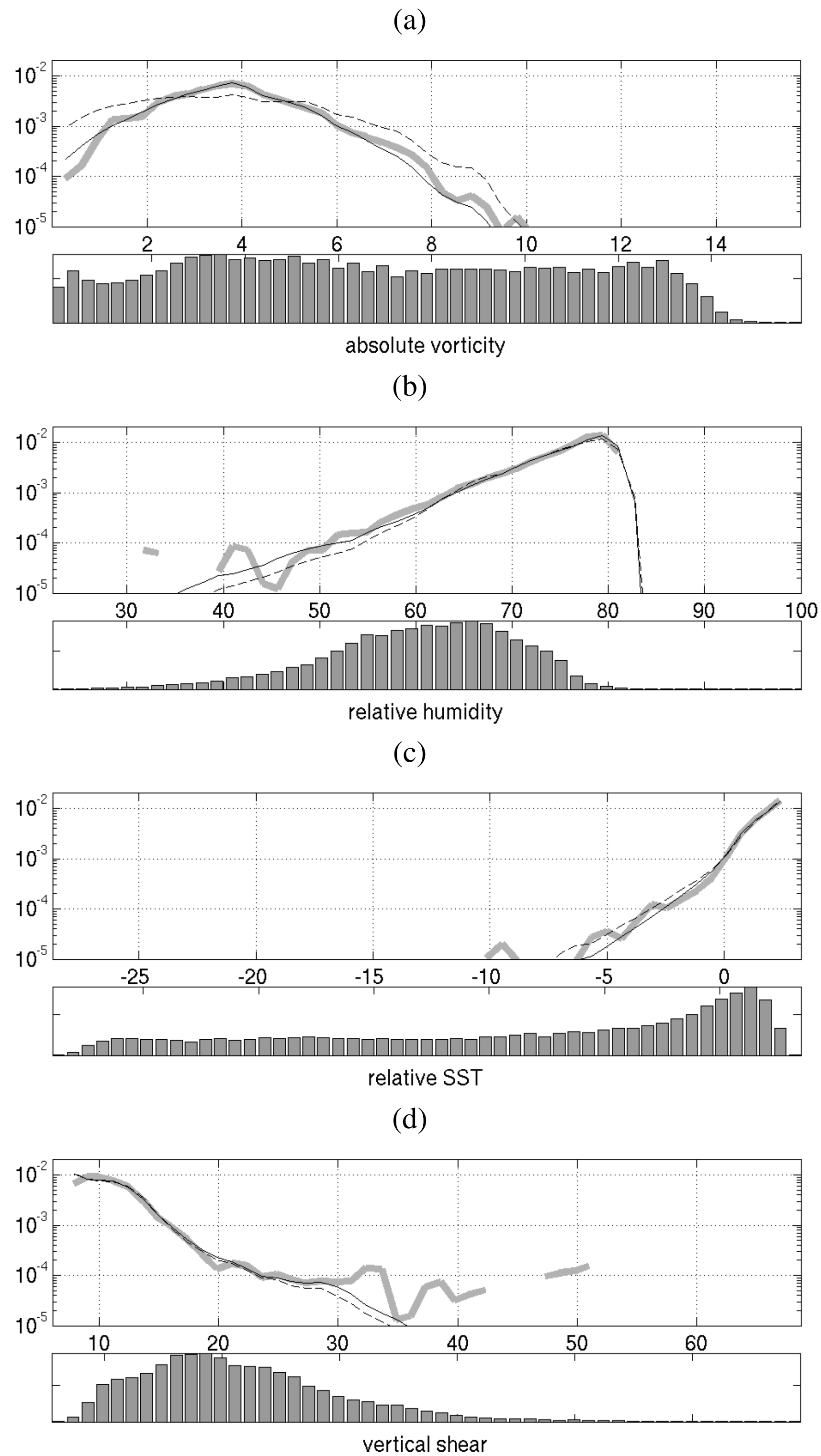

FIG. 11. Number of observed (thick line) and modeled (thin lines) tropical cyclones per year as a function of (a) absolute vorticity, (b) relative humidity, (c) relative SST and (d) vertical shear. The solid (dashed) thin line is the model with $\min(\eta, 3.7)$ (η) as a predictor. Histograms show the distribution of values of the climate variables.

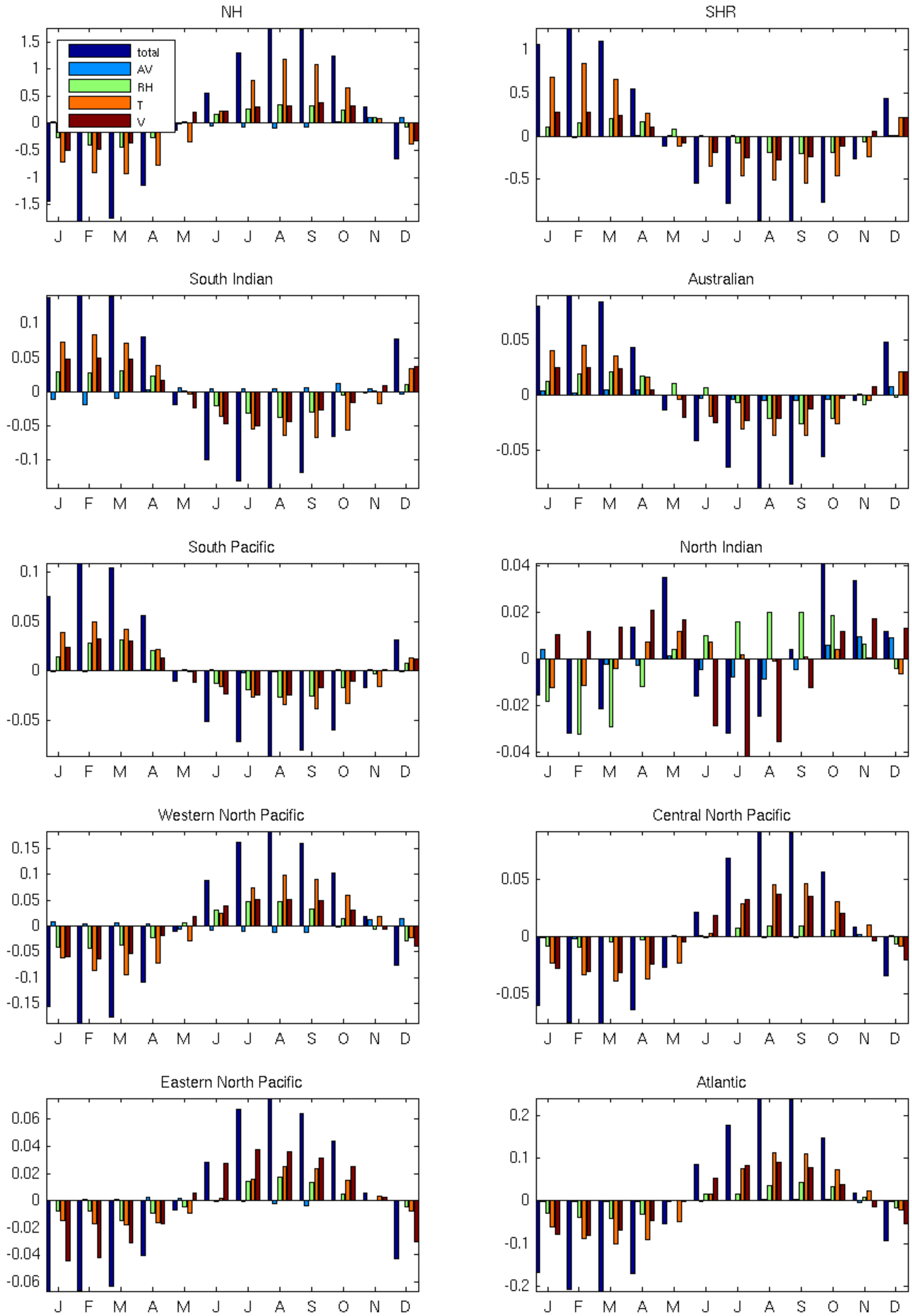

FIG. 12. Anomalies of the contributions of the clipped absolute vorticity (AV), relative humidity (RH), relative SST (T) and vertical shear (V) to the total.

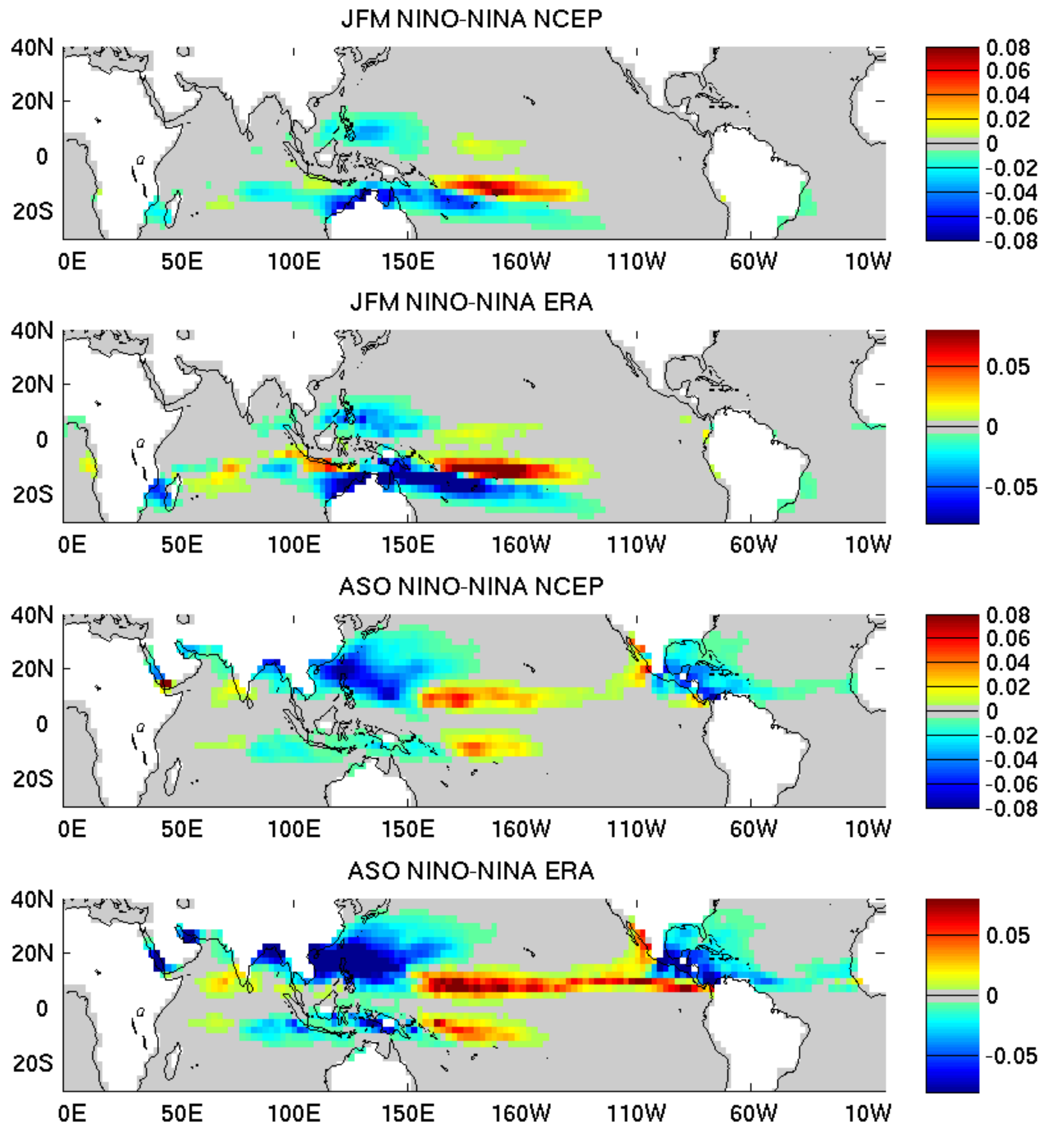

FIG. 13. Differences between El Niño and La Niña TCG index composites.