

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, and Department of Earth and Environmental Sciences, Columbia University, New York, New York

(Manuscript received 28 April 2010, in final form 28 September 2010)

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 versus potential intensity and satellite-based column-integrated relative humidity versus re-analysis relative humidity at a single level; these choices lead to modest differences in the performance of the index. The feature of the new index that 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. 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 because of 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 because of 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 the

distribution of tropical cyclone genesis 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 that 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

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, NY 10964.

E-mail: tippett@iri.columbia.edu

provide understanding. Empirical study of the problem using the observational record remains relevant.

Gray (1979) first developed an empirical “index” for tropical cyclogenesis. Gray’s index, and those that have been developed later, 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 modified 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 follows the same general 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 genesis potential index (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 (RH) at a single level as the other thermodynamic predictor, though reasonable, is not precisely justified and in practice generally requires use of assimilated humidity fields that 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 nonnegligible

probability of genesis, and in some regions of particular interest, such as the tropical Atlantic Main Development Region, the GPI underpredicts 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 that 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 because of the requirements of some particular application, to make use of new observations of the predictors (or different predictors) or for other unforeseen reasons.

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., Gadgil et al. 1984; Graham and Barnett 1987; Fu et al. 1990; Zhang 1993; Fu et al. 1994; Back and Bretherton 2009), our current understanding is that this threshold should not be fixed but should vary with the mean climate (whether because of anthropogenic or natural causes) because stability arguments show that the threshold is a function of tropical tropospheric temperature (e.g., Sobel et al. 2002; Chiang and Sobel 2002; Su et al. 2003). In particular, Yoshimura et al. (2006) and Knutson et al. (2008) show that the SST threshold for tropical cyclone formation rises in global warming simulations with high-resolution models. 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 increases the difficulty of computing the index and understanding 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(T)—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 (Ramsay and Sobel 2011). Relative SST is similar to Gray's original SST parameter but is a linear function of SST (no Heaviside function) and allows for change in the mean climate by using tropical mean SST in place of a fixed threshold 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 comparative 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. Since 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 that 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 climatological absolute vorticity (AV) is nonlinear. When the climatological absolute vorticity exceeds a threshold, further increase does not increase the climatological 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; and 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 $2.5^\circ \times 2.5^\circ$ longitude–latitude grid extending from 60°S to 60°N . Values of 850-hPa absolute vorticity, 600-hPa relative humidity, and vertical shear (V) between the 850- 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 et al. 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) datasets. Climatological means of the variables for both datasets were computed using the common 40-yr 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 water vapor W from the Remote Sensing Systems Inc. (see <http://www.remss.com>) for all available Special Sensor Microwave Imager (SSM/I; satellites F08, F10, F11, F12, F14, and F15) in the period 1987–2008. Details of the SSM/I retrieval algorithms are given in Wentz and Spencer (1998). The data are provided on a $0.25^\circ \times 0.25^\circ$ 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 $2.5^\circ \times 2.5^\circ$ grid for the region 60°S – 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 SSM/I data were available (NCEP/NCEP: 1987–2008, ERA-40: 1987–August 2002). The daily column-relative humidity is then defined as the ratio W/W_* . Then monthly means and climatological values are 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 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 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), and Camargo et al. (2009). The genesis potential index is defined as

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

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

The tropical cyclone genesis data comes from release v02r01 of the International Best Track Archive for Climate Stewardship (IBTrACS; Knapp et al. 2010) for the period 1961–2000. We define genesis locations as the first positions of those storms that eventually reach maximum sustained wind speed of $15 m s^{-1}$. Storms without maximum sustained wind speed data are not included. The exclusion of storms with missing maximum sustained wind speed data has the largest impact on the North Indian Basin.

b. Poisson regression

Poisson regression is typically used 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 $n = 0, 1, 2, \dots$, with probability

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

Here, for each $2.5^\circ \times 2.5^\circ$ grid cell and calendar month, N is the number of TCG events during a 40-yr 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,

$$\log \mu = \mathbf{b}^T \mathbf{x}, \quad (3)$$

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

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

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

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

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 the number of TCG events per area.

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

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

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. However, fitting both NCEP and ERA data means the observations in (6) are obviously not independent. Therefore, we deflate the number of observations by a factor of 2. This scaling reduces the log likelihood and increases the standard errors of the coefficient estimates.

The maximized log likelihood is one 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 as

$$AIC = -2L + 2p, \quad (7)$$

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.

The variance of a Poisson distribution is equal to its mean. In practice, data often exhibits greater variability, a property denoted as overdispersion. The dispersion

parameter σ measures the ratio of variance to mean with $\sigma > 1$ corresponding to overdispersion. The dispersion parameter σ is estimated from

$$\sigma^2 = \frac{1}{k-p} \sum_{i=1}^k \frac{(N_i - \mu_i)^2}{\mu_i}. \quad (8)$$

A common way to deal with overdispersion, and the approach taken here, is the so-called quasi-Poisson method, in which the coefficient estimates are same as in Poisson regression, but their standard errors are inflated to reflect the overdispersion.

3. Construction of a TCG index

Our goal is to construct an index that 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. Here, we consider the following:

- Is reanalysis relative humidity adequate given that there are relatively few humidity measurements 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.

First, we consider the Poisson regression based on four climate variables: absolute vorticity, reanalysis midlevel relative humidity, relative SST, and vertical shear. The simplest Poisson regression model assumes no interactions between the predictor variables, that is, no powers or products of the predictors are included in the model, and it has the form

$$\mu = \exp(b + b_\eta \eta + b_{\mathcal{H}} \mathcal{H} + b_T T + b_V V + \log \cos \phi), \quad (9)$$

where μ is the expected number of tropical cyclone genesis events per month in a 40-yr period, 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 $^\circ\text{C}$, and vertical shear between the 850- and 200-hPa levels in m s^{-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 line 1 of the Table 1 estimates. Standard errors for the coefficient estimates are computed from regression statistics and from 1000 bootstrap samples and are shown in Table 1, confirming the significance of the coefficients.

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 expected number of TCG events is

$$\frac{\delta \mu}{\mu} \approx \mathbf{b}^T \delta \mathbf{x}. \quad (10)$$

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 index.

Addressing the first question regarding the choice of humidity variable, we find that using the SSM/I column-integrated relative humidity rather than the reanalysis relative humidity gives a lower AIC value (line 2 of Table 1), indicating a better fit to observations. However, fitting the NCEP and ERA data separately reveals that using the SSM/I 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 does not facilitate the use of a single Poisson regression model. The hemisphere- and basin-averaged column-integrated relative humidities computed using SSM/I water vapor path divided by saturation values derived from NCEP and ERA temperatures are more similar to each other (not shown), indicating better agreement in temperature than humidity, as might be expected.

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 the purpose of fitting the spatial distribution of genesis probability from the present climatology, SST contains nearly the same information as relative SST; they differ only by the annual cycle of mean tropical SST. As the climate varies (because of either anthropogenic or natural causes), we expect the

TABLE 1. Estimate of the 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 index.

	Vorticity			Humidity		Heat		Shear	AIC	σ^2	
	b	b_{η}	b_{η^2}	$b_{\min(\eta,3.7)}$	b_H	$b_{H_{SSM/I}}$	b_T	b_{PI}	b_V		
Estimates											
1	-4.47	0.50	—	—	0.05	—	0.63	—	-0.17	13208.32	3.46
2	-11.15	0.56	—	—	—	0.13	0.57	—	-0.15	12895.37	2.76
3	-16.25	0.60	—	—	—	0.08	—	0.19	-0.14	13685.06	5.05
4	-12.73	1.80	-0.16	—	—	0.12	0.43	—	-0.14	12316.44	13.75
5	-11.96	—	—	1.12	—	0.12	0.46	—	-0.13	12213.35	2.90
6	-5.80	—	—	1.03	0.05	—	0.56	—	-0.15	12486.84	5.00
Standard errors (regression statistics)											
1	0.25	0.021	—	—	0.0034	—	0.035	—	0.01		
2	0.54	0.02	—	—	—	0.0069	0.03	—	0.01		
3	0.72	0.029	—	—	—	0.0053	—	0.0082	0.02		
4	1.2	0.23	0.03	—	—	0.015	0.085	—	0.02		
5	0.54	—	—	0.044	—	0.0067	0.037	—	0.01		
6	0.33	—	—	0.055	0.004	—	0.047	—	0.01		
Standard errors (bootstrap statistics)											
1	0.19	0.011	—	—	0.0033	—	0.016	—	0.01		
2	0.44	0.013	—	—	—	0.0055	0.016	—	0.01		
3	0.5	0.015	—	—	—	0.0046	—	0.0041	0.01		
4	0.41	0.086	0.012	—	—	0.0051	0.028	—	0.01		
5	0.37	—	—	0.031	—	0.0046	0.024	—	0.01		
6	0.19	—	—	0.026	0.0031	—	0.021	—	0.01		

threshold for deep convection to vary roughly with the tropical mean SST (Sobel et al. 2002, e.g.,) 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 a highly derived quantity whose computation, compared to relative SST, is 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 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 four climate variables one at a time

to the Poisson regression, we find that including the square of the absolute vorticity reduces AIC the most (line 4 of Table 1). The resulting negative coefficient (-0.16) for the square of absolute vorticity (line 4 of Table 1) means that in this index, 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

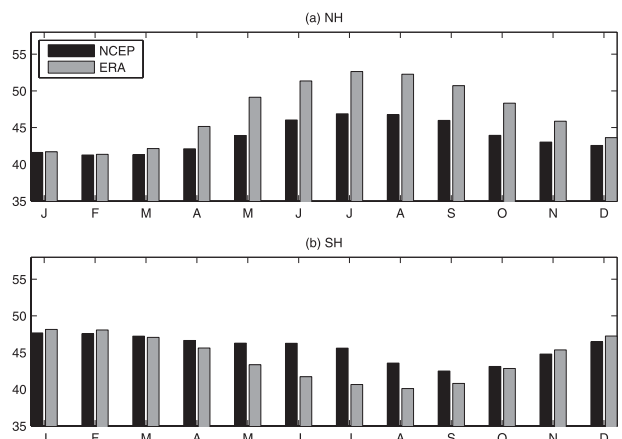


FIG. 1. Hemispheric (ocean grid points) averaged 600-hPa relative humidity from the NCEP reanalysis and ERA: (a) NH and (b) SH.

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, we scale ERA PI by the factor 0.72.

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 and make the estimation of its parameters less robust. Note also, there is a substantial increase in the dispersion parameter σ^2 (line 4 of Table 1).

To examine further 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 errors inflated by 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 $4 \times 10^{-5} \text{ s}^{-1}$, 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 (light 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 minimum ($\eta, 3.7$) 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

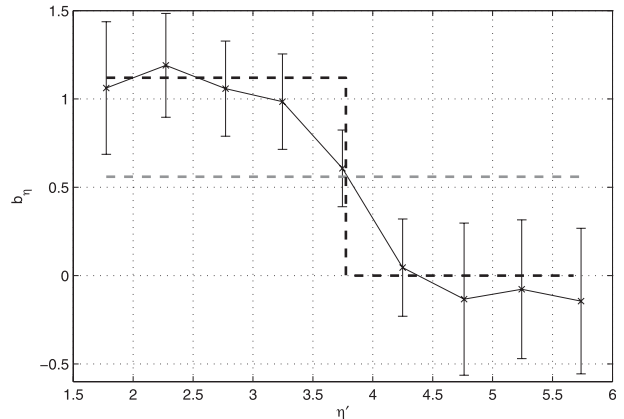


FIG. 2. The solid line shows b_η as a function of η' fit using values of the absolute vorticity in the interval $(\eta' - 0.5, \eta' + 0.5)$; error bars represent the 95% confidence intervals. The light dashed line is the value of b_η obtained using all the data. The dark dashed line corresponds to using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

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 minimum ($\eta, 3.7$). The coefficient of $\min(\eta, 3.7)$ is 1.12 which is larger than that of η (0.56; line 2 of Table 1). We will see later that this feature means that the index based on clipped absolute vorticity responds more strongly to the near-equatorial latitudinal gradient 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 substantially reduce AIC or change the fitted values. 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 Fig. 3. For the most part, there is reasonable agreement, both in spatial structure and in magnitude, between the observations and the indices. However, the magnitude of the NCEP TCG index is weak in the Atlantic main development region. The region of maximum observed density of TCG events in the Northern Pacific extends farther equatorward than is seen in either of the indices. Neither index matches

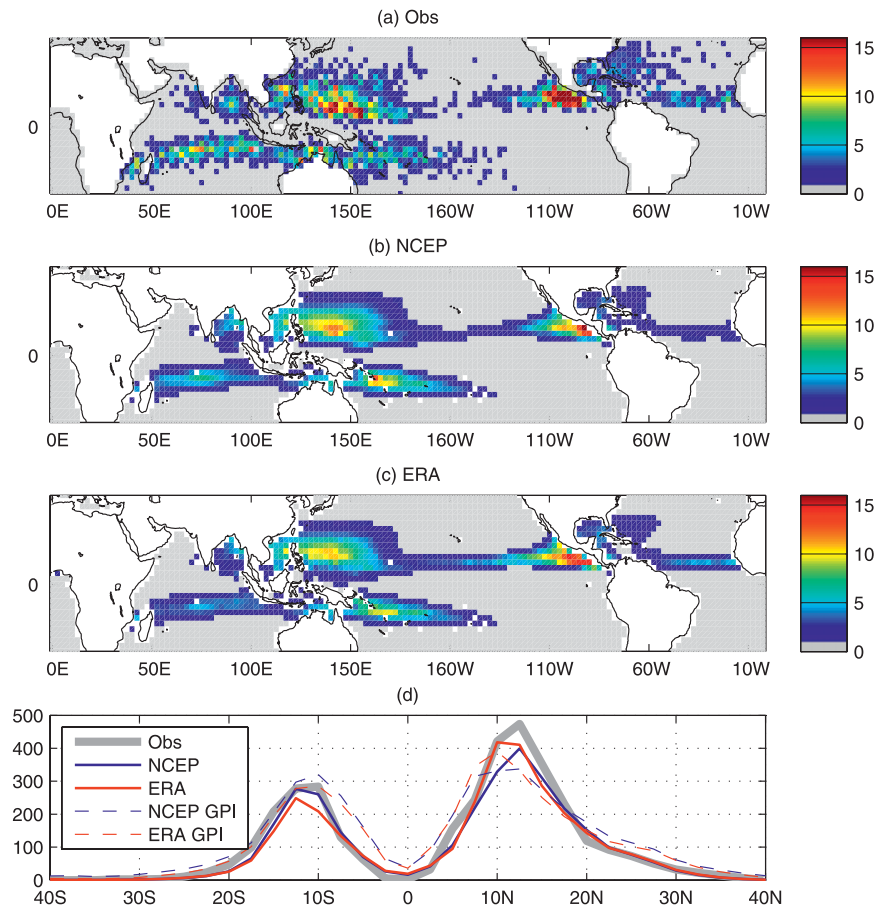


FIG. 3. Number of (a) observed TCG events over the 40-yr climatology period and TCG indices based in (b) NCEP and (c) ERA data. (d) Zonal sums of the observations and indices.

the observed values in the Arabian Sea and the southern part of the North Indian Ocean. Values of the indices in the South Pacific are too large compared to observations while values in the eastern Pacific are too weak. Zonal sums of the observations and indices in Fig. 3d show that the NCEP TCG index, and to a lesser extent the ERA TCG index, is too weak in the Northern Hemisphere. The agreement in the Southern Hemisphere between the zonal sums of the observations and indices masks what we show later are off-setting biases in the annual cycle. Both indices show small but nonzero likelihood of TCG on the equator. For comparison we also show zonal sums of the GPI based on NCEP and ERA data.

We examine the seasonal cycle of the TCG index by comparing the basin-integrated climatology of the TCG index with that of the observations; basin 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

TABLE 2. Domain definitions used for basin integrations.

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

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 the inactive austral winter period. The two peaks in the North Indian seasonal cycle are reproduced. 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 observations during the active phase 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.

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 datasets. 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 the worst in the western North Pacific peak season. Figures 5g,h and 6g,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 because of it 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 owing to the reduction in the South Indian and Australian Basins. Figures 5i,j and 6i,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 for ASO 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 latitudinally.

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 systematically different amplitudes and, therefore, so do the GPI values that 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 (in the sense of minimizing the sum of squared errors) the observed number of TCG events. Figure 10 shows that overall GPI is too high in the Southern Hemisphere especially during the inactive season. The GPI peak occurs about a month later in the South Indian basin. Amplitudes of the GPI in the western North Pacific and Atlantic match observations well during the active season but are too strong during the inactive months. Presumably, the imposed scaling of the GPI leads to better fitting in the Northern Hemisphere (where the total number of storms is larger) at the expense of fitting the Southern Hemisphere. Given the somewhat arbitrary scaling of the GPI, it is difficult to say whether the GPI is too small in the Northern Hemisphere or too large in the Southern Hemisphere. A more precise statement is that the difference in the number of storms in the Northern and Southern Hemispheres is too small in the GPI. Figures 5k,l and 6k,l show that GPI spatial patterns extend too far poleward, are too close to the equator (features that are consistent with Fig. 3d), and show less difference between JFM and ASO amplitudes.

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

$$N_{\eta}(\eta') = \langle N(\eta_i, \mathcal{H}_i, T_i, V_i) \rangle, \quad \text{where} \\ i \text{ satisfies: } (\eta' - \delta) \leq \eta_i \leq (\eta' + \delta); \quad (11)$$

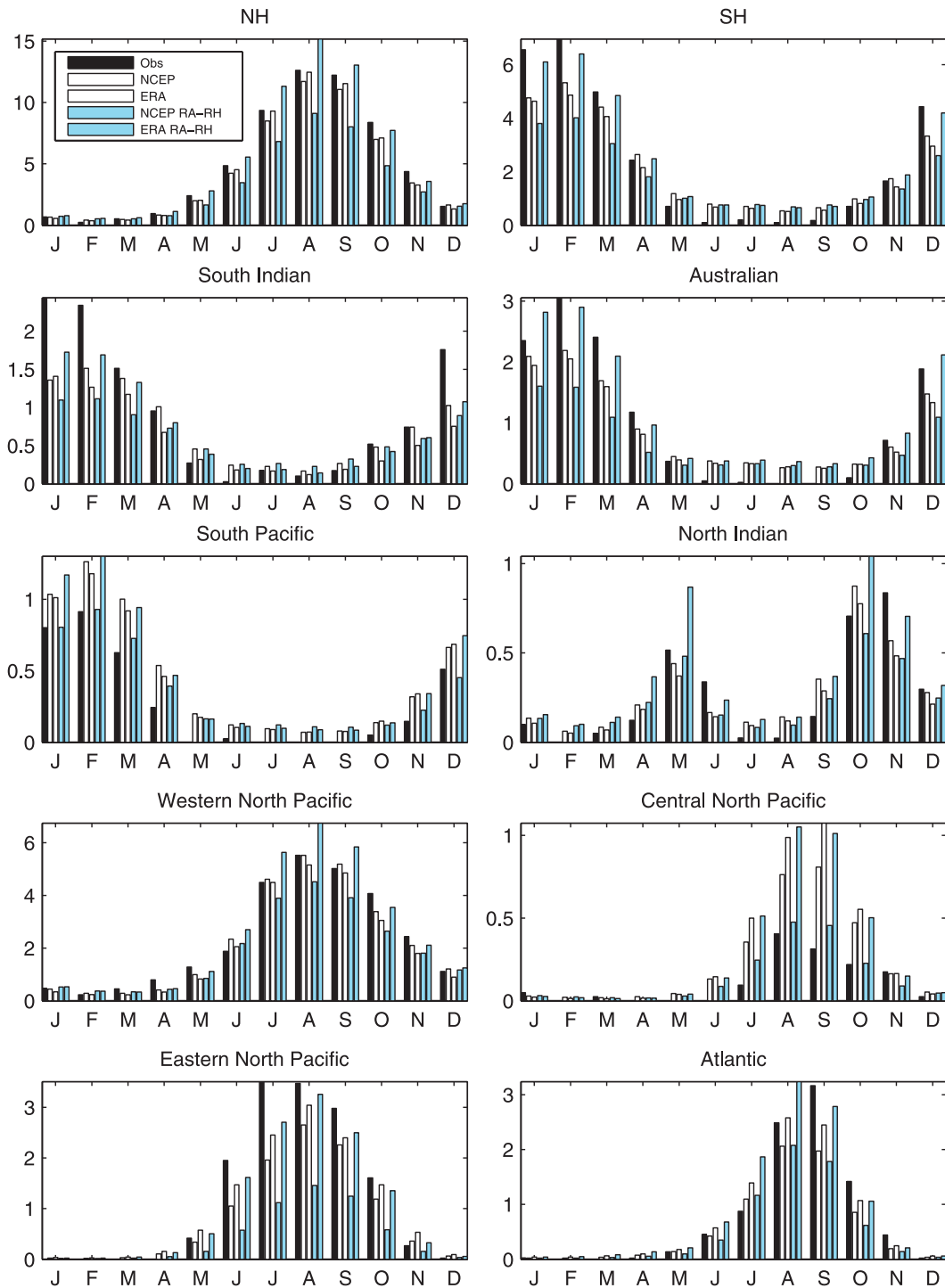


FIG. 4. Basin-integrated climatology (number of TCG events per year). The term RA-RH indicates results from the Poisson regression model with reanalysis 600-hPa relative humidity rather than SSM/I column-integrated relative humidity.

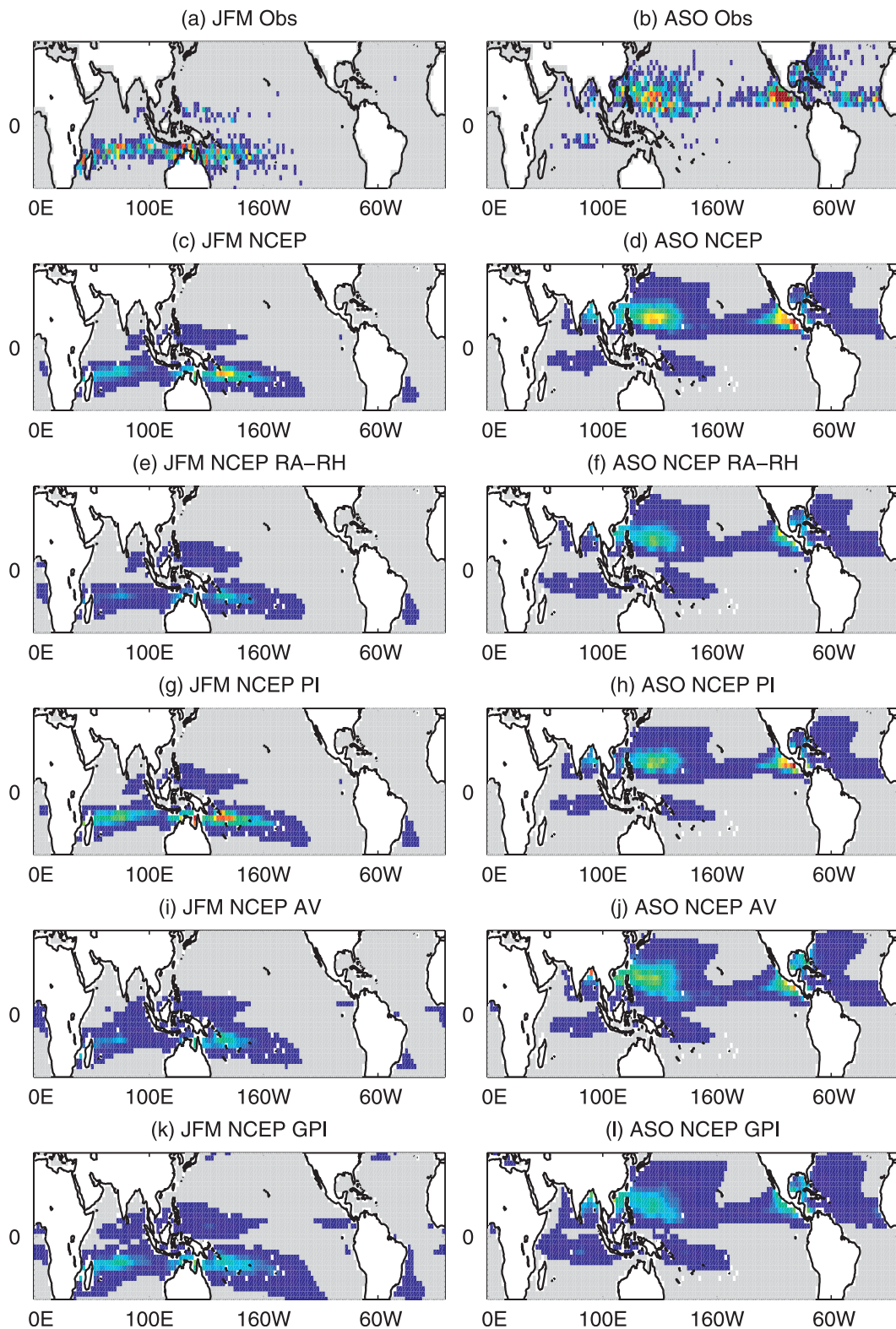


FIG. 5. Observed and modeled TCG peak season maps based on NCEP data. The term RA-RH indicates use of the reanalysis relative humidity; PI indicates use of PI rather than relative SST; AV indicates use of absolute vorticity rather than clipped absolute vorticity; GPI indicates use of GPI. Color scale as in Fig. 3 with maximum value of 10.

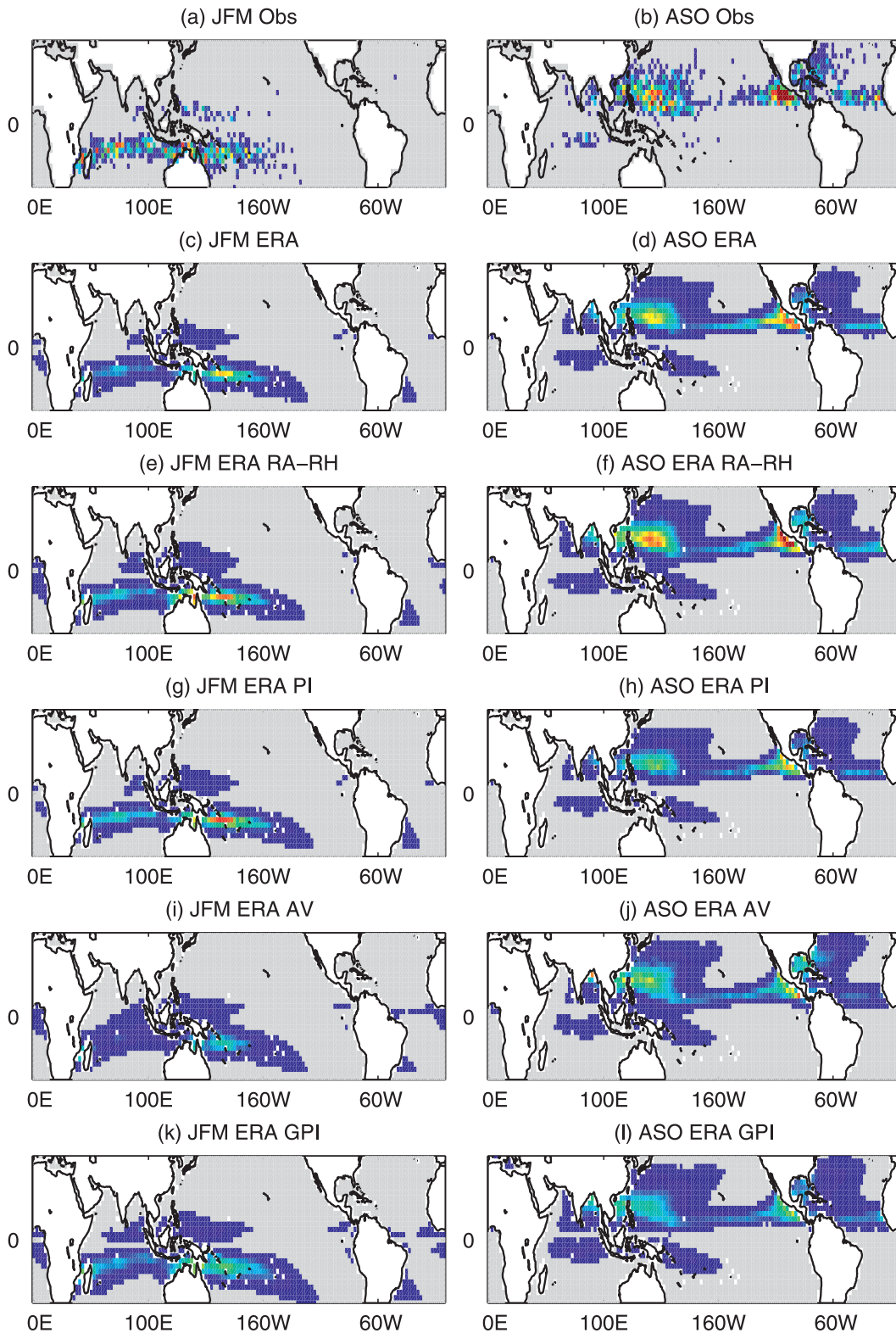


FIG. 6. As in Fig. 5, but for ERA data.

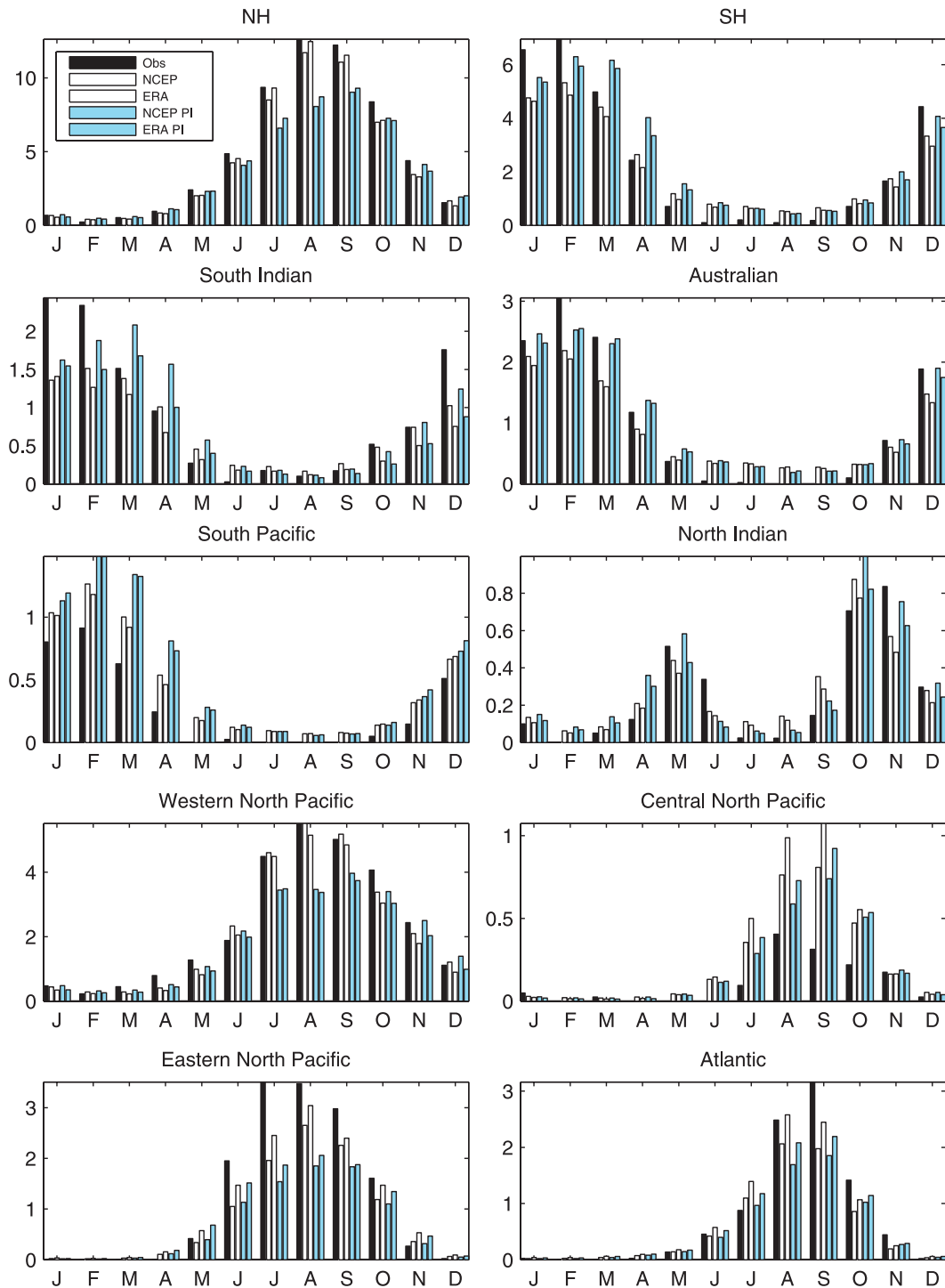


FIG. 7. As in Fig. 4, but with PI indicating the results from the Poisson regression model with PI rather than relative SST.

$\langle \cdot \rangle$ denotes average. Analogous marginal functions can be defined for relative humidity, relative SST, and vertical shear. The number of TCG events has a log-linear dependence on the climate variables in the Poisson regression model. However, the dependence of the marginal

function on the individual variables may not be log linear because of the correlations between the climate variables. Most of the correlations and hence much of the behavior of the marginal functions can be inferred from the latitude dependence of the zonally averaged climate variables.

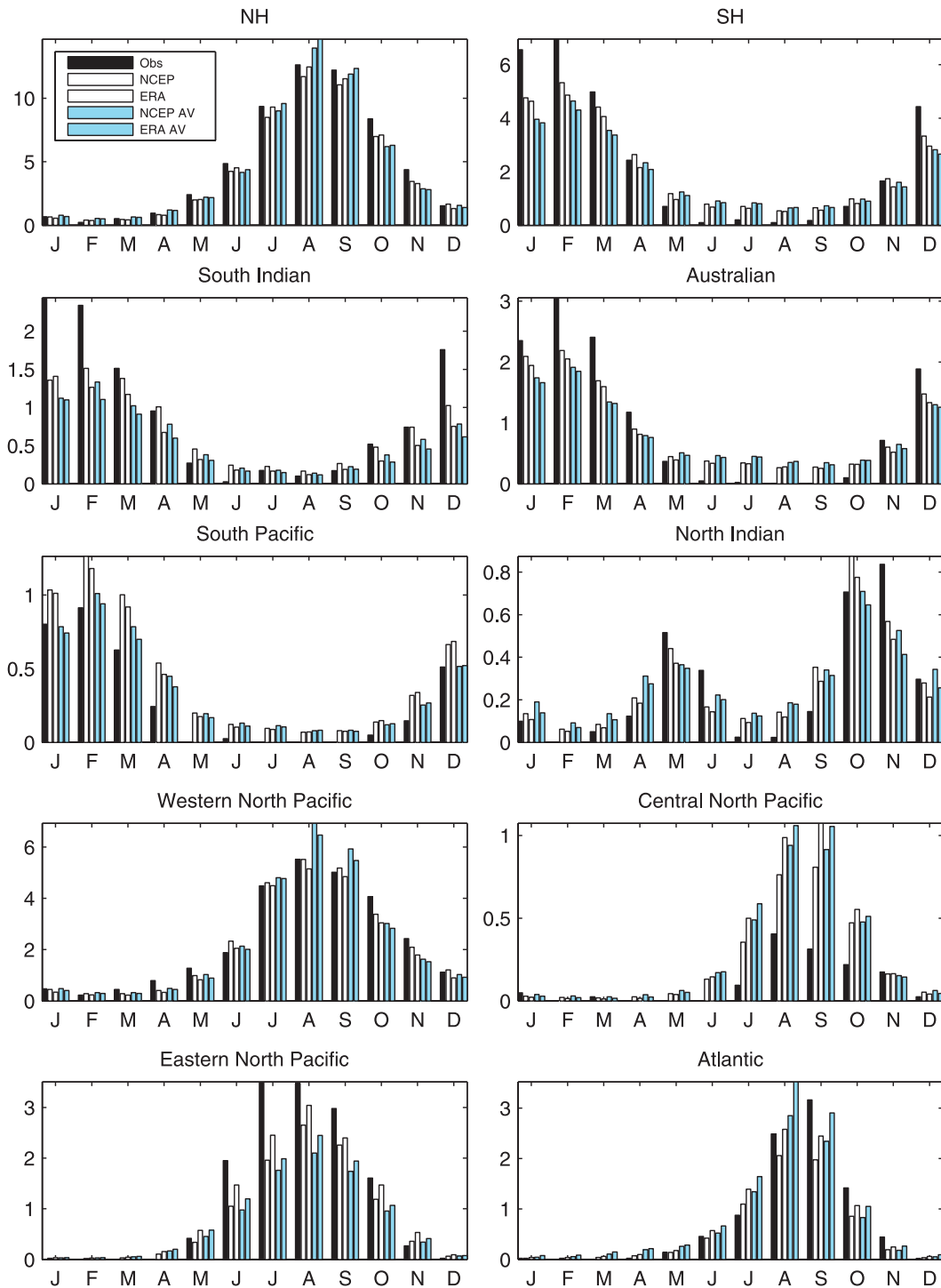


FIG. 8. As in Fig. 4, but with AV indicating the results from the Poisson regression model with absolute vorticity rather than clipped absolute vorticity.

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

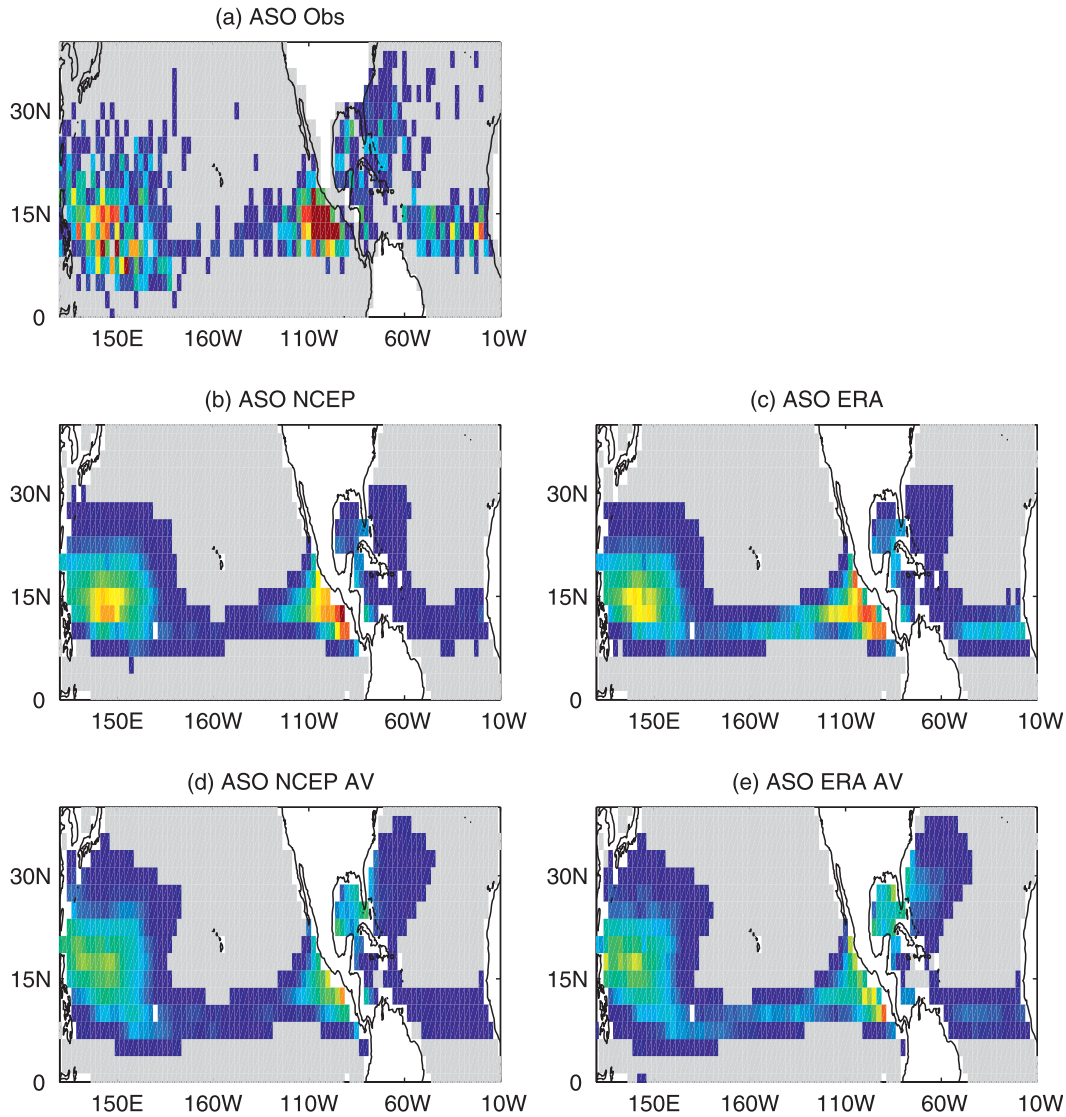


FIG. 9. Zoom in of Figs. 5i,j and 6i,j. The term AV is as in Fig. 8.

primarily in 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 in Fig. 11a shows the behavior for a model based on absolute vorticity rather than clipped absolute vorticity. Not using clipped absolute vorticity leads to an index that 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

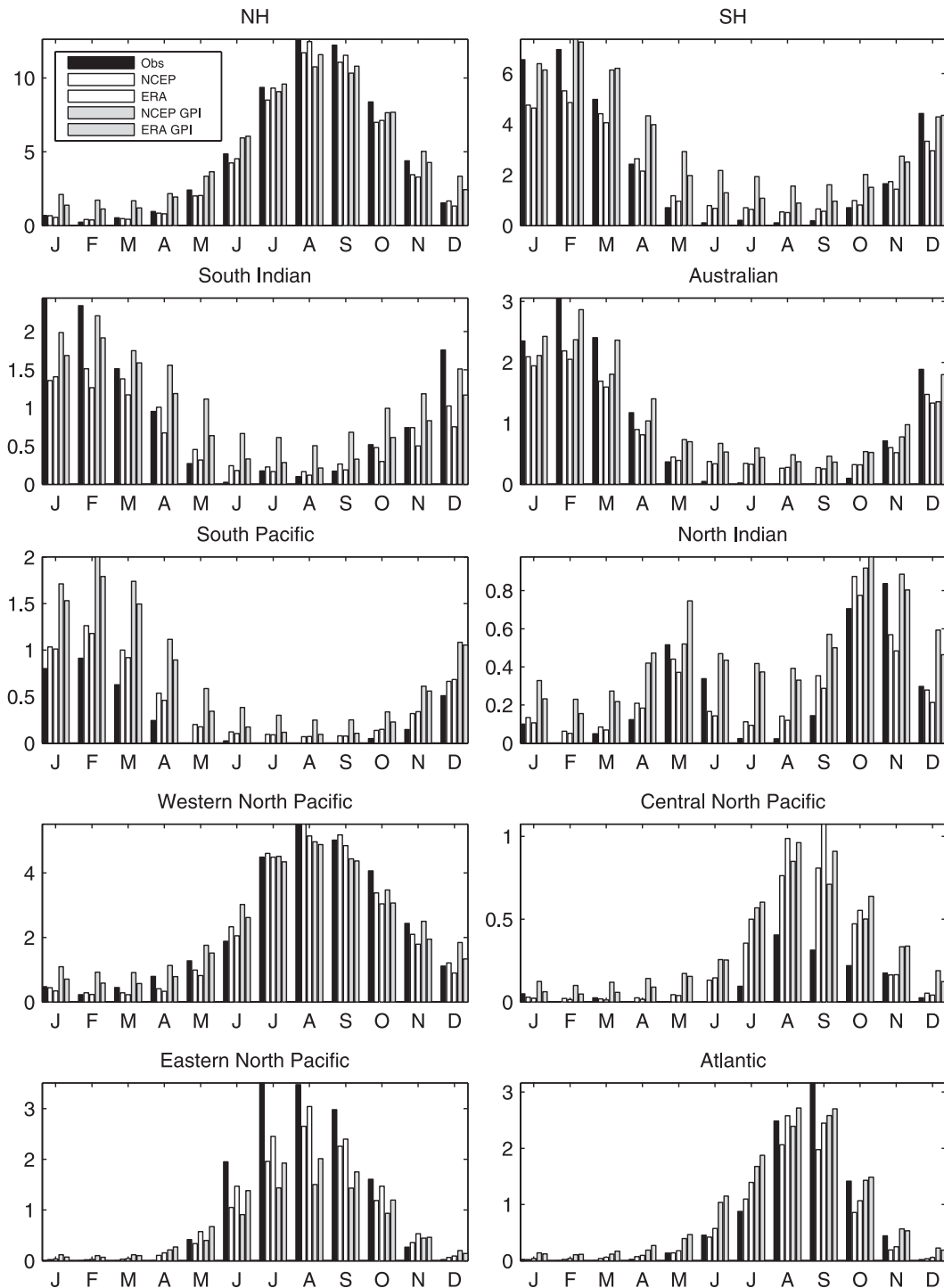


FIG. 10. As in Fig. 4, but with GPI indicating the area-weighted and scaled GPI.

equator where absolute vorticity is small and there are few TCG events.

For very small values of vertical shear, the index marginal function for vertical shear decreases little as vertical shear increases, and the observed number of

storms actually increases slightly (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 conditions that

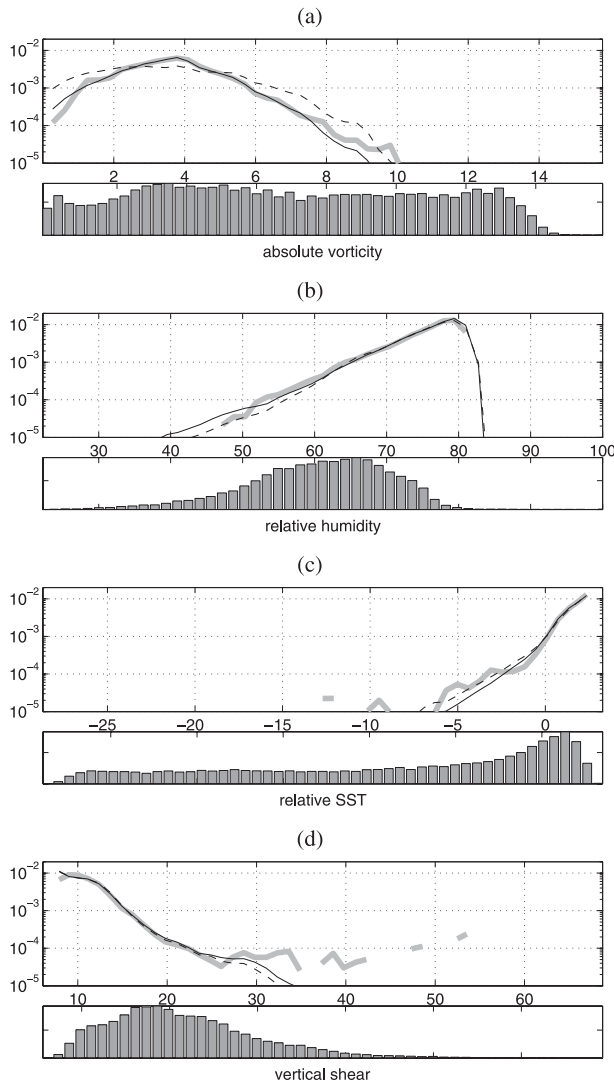


FIG. 11. Number of observed (thick line) and modeled (thin lines) tropical cyclones per year as a function of (a) absolute vorticity, (b) relative humidity, (c) relative SST, and (d) vertical shear. The solid (dashed) thin line is the model with the minimum (η , 3.7) (η) as a predictor. Histograms show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

are more favorable for TCG. 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), presumably in many cases deriving energy from the shear via baroclinic eddy dynamics (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; anomalies are with respect to the annual average. 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 cycles of the three factors are in phase. However, the behavior in the North Indian Basin with its two maxima is more complex with relative SST playing a minor role compared to vertical shear (Gray 1968; Evan and Camargo 2011). There, the reduction in vertical shear results in the premonsoon maximum. The increase in relative humidity following the start of the monsoon is offset by increases in vertical shear and the number of storms decreases. The reduction in vertical shear and continuing high relative humidity results in the second maximum.

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; Tables 5 and 6 show root-mean-squared (RMS) errors. 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 calculations 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 19 season basins with significant correlations while the ERA-based index has 15. For the most part, the TCG indices based on the two datasets show similar correlation levels and seasonality. Neither the NCEP nor the ERA index shows any significant interannual correlation in the North Indian basin. However, NCEP has some significant

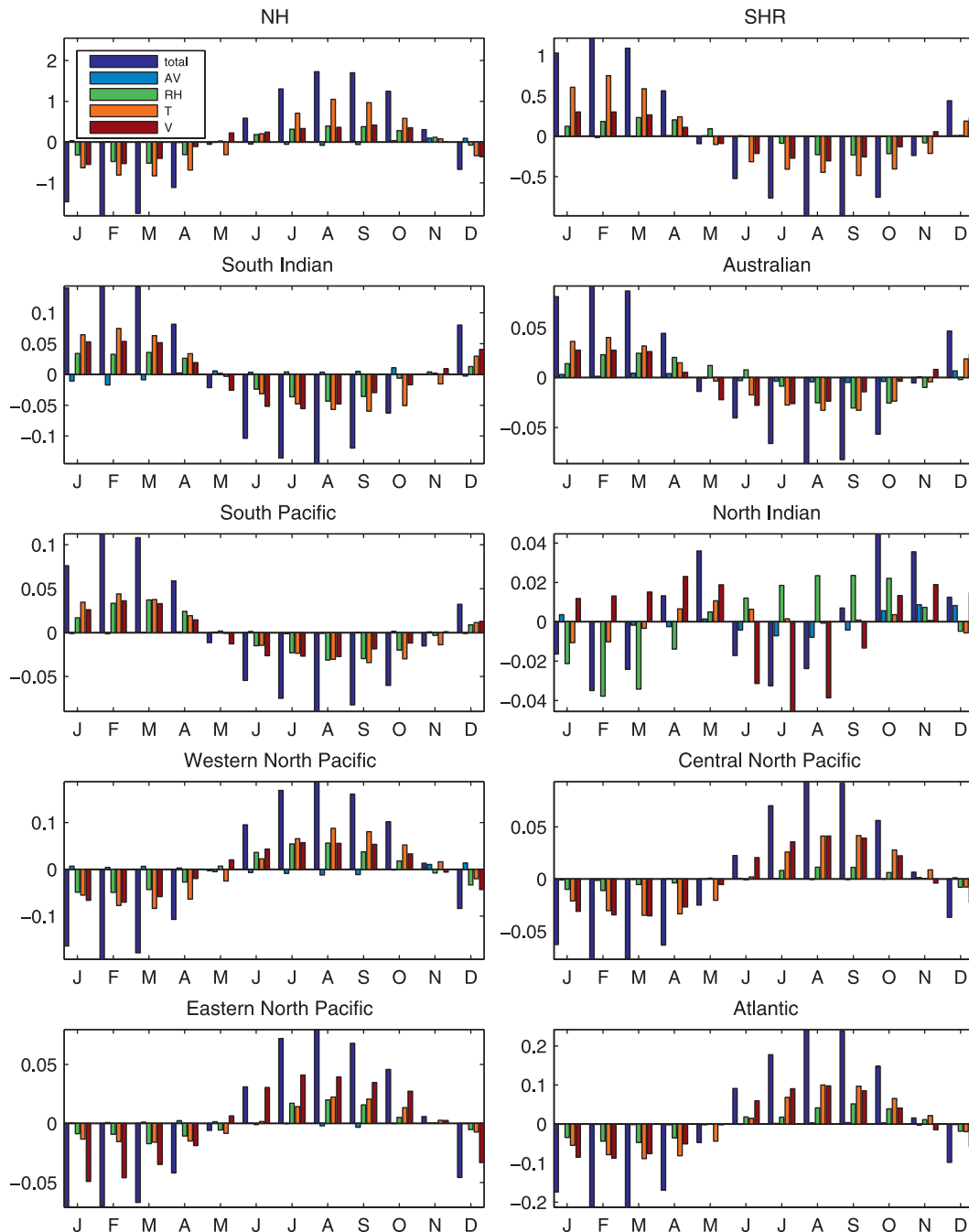


FIG. 12. Anomalies of the contributions of the clipped AV, RH, T , and V to the log of the TCG index.

correlations in the Australian region during austral summer while ERA does not. Correlation levels are roughly comparable to those found in Camargo et al. (2007a) using GPI and different observed genesis data. The RMS errors of the ERA-based index are considerably larger than those of the NCEP-based index, especially in the WNP during boreal summer. The large errors in the ERA-based index can be attributed to the behavior of

the ERA 600-hPa relative humidity; Daoud et al. (2009) found large deviations in 850-hPa relative humidity data from NCEP and ERA over the North Atlantic Ocean. Figure 13 shows the NCEP and ERA July–September (JAS) 600-hPa relative humidity averaged over the box $120^{\circ}\text{E}–180^{\circ}$, $0^{\circ}–30^{\circ}\text{N}$. Prior to 1972 the two analyses agree. After 1972, the ERA relative humidity exhibits greater variability and exceeds its 1961–2000 climatological value.

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 > 0.4 are in bold.

	SI	Aus	SP	NI	WNP	CNP	ENP	Atl
JFM	-0.14	0.46	0.43	—	—	—	—	—
FMA	0.15	0.5	0.49	—	0.4	—	—	—
MAM	-0.047	0.38	0.69	-0.04	0.39	—	—	—
AMJ	-0.051	0.13	—	-0.13	0.22	—	0.23	—
MJJ	—	—	—	-0.16	0.52	—	0.072	0.075
JJA	—	—	—	—	0.36	—	0.11	0.27
JAS	—	—	—	—	0.097	0.45	0.33	0.59
ASO	—	—	—	0.0077	0.0068	0.69	0.54	0.56
SON	0.22	—	—	-0.34	0.046	0.85	0.68	0.38
OND	0.065	0.29	—	-0.22	0.28	—	0.39	0.4
NDJ	0.22	0.44	0.7	-0.035	0.22	—	—	—
DJF	-0.016	0.34	0.5	—	0.18	—	—	—

Consequently, during the period 1982–2001, the ERA-based index presents values considerably larger than its 1961–2000 climatological value.

Much of the interannual variability is related to El Niño–Southern Oscillation (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 Niño-3.4 index were classified as El Niño years and the 11 years with the lowest values as La Niña years. Table 7 shows the years selected for the composites. Figure 14 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 equatorward 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

TABLE 5. RMS difference between seasonal basin-integrated observed number of TCG events and TCG index based on NCEP data. Seasons averaging <1 TCG event per year are not included.

	SI	Aus	SP	NI	WNP	CNP	ENP	Atl
JFM	3.3	3.1	2.3	—	—	—	—	—
FMA	2.4	3.1	2.2	—	1.4	—	—	—
MAM	1.7	2	1.5	0.9	1.6	—	—	—
AMJ	0.92	1.2	—	1.2	2.6	—	1.5	—
MJJ	—	—	—	1.4	2.9	—	3.6	2
JJA	—	—	—	—	3.3	—	4.6	2.8
JAS	—	—	—	—	3.7	1.1	4.8	2.6
ASO	—	—	—	1.7	3.7	0.95	3.4	3.2
SON	1.1	—	—	1.3	3.2	0.81	2.5	3
OND	1.8	1.7	—	1.4	2.7	—	1.4	1.9
NDJ	2.5	2.3	1.8	1.1	2.5	—	—	—
DJF	3.3	2.9	1.9	—	2.3	—	—	—

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. The precise dependence of the probability of genesis on the value of the background vorticity is less clear. 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

TABLE 4. As in Table 3, but for ERA data.

	SI	Aus	SP	NI	WNP	CNP	ENP	Atl
JFM	-0.45	0.26	0.38	—	—	—	—	—
FMA	0.19	0.26	0.49	—	0.44	—	—	—
MAM	-0.13	0.16	0.76	0.14	0.38	—	—	—
AMJ	-0.15	-0.091	—	0.21	0.14	—	0.011	—
MJJ	—	—	—	0.19	0.62	—	0.059	0.42
JJA	—	—	—	—	0.34	—	0.14	0.6
JAS	—	—	—	—	0.43	0.29	0.31	0.57
ASO	—	—	—	-0.016	0.12	0.63	0.38	0.54
SON	0.19	—	—	-0.31	0.19	0.82	0.42	0.28
OND	0.087	0.21	—	-0.36	0.19	—	0.29	0.43
NDJ	-0.11	0.35	0.68	-0.17	0.095	—	—	—
DJF	-0.36	0.11	0.33	—	0.025	—	—	—

TABLE 6. As in Table 5, but for ERA data.

	SI	Aus	SP	NI	WNP	CNP	ENP	Atl
JFM	3.2	9.4	4.1	—	—	—	—	—
FMA	2.3	6.8	3.6	—	2	—	—	—
MAM	2.4	5.4	2.5	1.9	3.6	—	—	—
AMJ	1.6	4.2	—	2.3	7.3	—	3	—
MJJ	—	—	—	2.4	12	—	4	3.5
JJA	—	—	—	—	17	—	6.2	5.8
JAS	—	—	—	—	18	3.3	6.4	6.7
ASO	—	—	—	3.8	16	3.3	5.7	5
SON	1.3	—	—	3.3	11	2.2	3.5	3.1
OND	2	6.5	—	2.6	7.4	—	2	1.5
NDJ	2.6	9.5	2.9	1.3	4.9	—	—	—
DJF	3	10	3.7	—	3.5	—	—	—

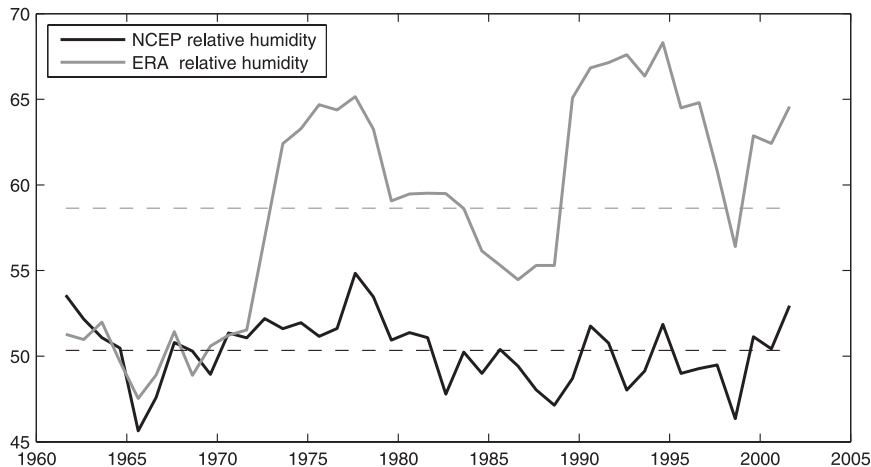


FIG. 13. The JAS 600-hPa relative humidity averaged over the box $120^{\circ}\text{E}\text{--}180^{\circ}$, $0^{\circ}\text{--}30^{\circ}\text{N}$ from the NCEP (dark) and ERA (gray) datasets. Dashed lines indicate the 1961–2000 average.

fields. On daily time scales, it is certain that the existence of a preexisting tropical depression makes genesis more likely compared to the absence of a depression, and it seems likely that the vorticity of the depression, which presumably in many cases exceeds $4 \times 10^{-5} \text{ s}^{-1}$, 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. If one were to use the index derived here—or any similar one derived from monthly or climatological data—for prediction on shorter time scales, it would be important to reconsider this issue and perhaps to modify the derivation of the index to account for the time scale dependence of the vorticity influence.

6. Summary and conclusions

The likelihood of tropical cyclone genesis (TCG) being observed depends on features of the large-scale climate. Therefore, changes in climate because of 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 that 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-yr 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 datasets. 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- 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

TABLE 7. Years used to form ENSO composites.

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

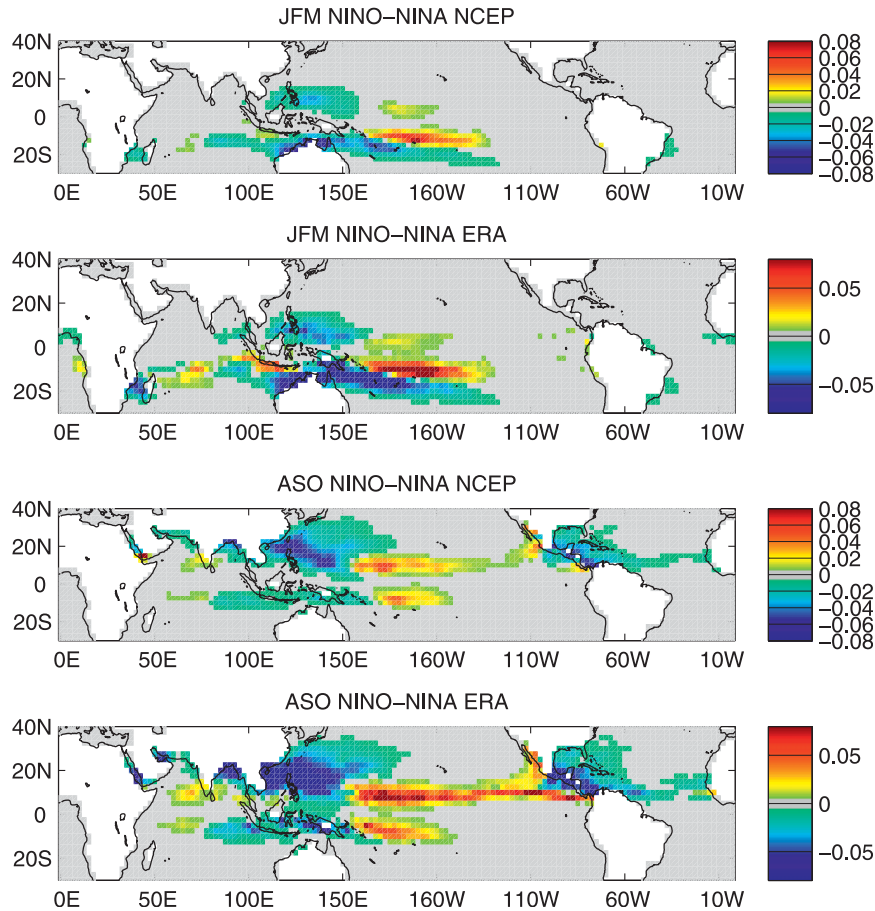


FIG. 14. Differences between El Niño and La Niña TCG index composites.

nonzero for values of absolute vorticity less than $4 \times 10^{-5} \text{ s}^{-1}$, while for values of absolute vorticity greater than $4 \times 10^{-5} \text{ s}^{-1}$, it is close to 0. 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 that 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 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 (Ramsay and Sobel 2011). 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 observed marginal dependence of the number of TCG events on the individual climate variables. This dependence is not log linear owing to the 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 is 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 covariability 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 because of ENSO. Future work will apply the index to simulated climate change scenarios.

Acknowledgments. We thank the anonymous reviews for their useful comments and suggestions. We thank Larissa Back and Chris Bretherton for making available the scripts for calculation of column-relative humidity. We especially thank Larissa Back for her help and explanations. We also thank Kerry Emanuel and Gabriel Vecchi for discussions. The ECMWF ERA-40 data used in this study were obtained from the ECMWF data server. SSM/I data are produced by Remote Sensing Systems and sponsored by the NASA Earth Science MEaSUREs DISCOVER Project. Data are available online at www.remss.com. 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 subagencies.

REFERENCES

- Akaike, H., 1973: Information theory and an extension of the maximum likelihood principle. *2nd International Symposium on Information Theory*, B. N. Petrov and F. Czaki, Eds., Akademiai Kiadó, 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*, **59A**, 396–416.
- Bister, M., and K. A. Emanuel, 1998: Dissipative heating and hurricane intensity. *Meteor. Atmos. Phys.*, **52**, 233–240.
- , and —, 2002a: Low frequency variability of tropical cyclone potential intensity. 1. Interannual to interdecadal variability. *J. Geophys. Res.*, **107**, 4801, doi:10.1029/2001JD000776.
- , and —, 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 diagnose ENSO effects on tropical cyclone genesis. *J. Climate*, **20**, 4819–4834.
- , A. H. Sobel, A. G. Barnston, and K. A. Emanuel, 2007b: Tropical cyclone genesis potential index in climate models. *Tellus*, **59A**, 428–443.
- , 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-scale environmental parameters associated with tropical cyclone formations in the western North Pacific. *J. Climate*, **17**, 466–484.
- Chiang, J. C. H., and A. H. Sobel, 2002: Tropical tropospheric temperature variations caused by ENSO and their influence on the remote tropical climate. *J. Climate*, **15**, 2616–2631.
- Daoud, A. B., E. Sauquet, M. Langand, C. Obled, and G. Bontron, 2009: Comparison of 850-hPa relative humidity between ERA-40 and NCEP/NCAR re-analyses: Detection of suspicious data in ERA-40. *Atmos. Sci. Lett.*, **10**, 43–47.
- 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. Schertmann, 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.
- , 1995: Sensitivity of tropical cyclones to surface exchange coefficients and a revised steady-state model incorporating eye dynamics. *J. Atmos. Sci.*, **52**, 3969–3976.
- , and D. S. Nolan, 2004: Tropical cyclone activity and global climate. *Proc. 26th Conf. on Hurricanes and Tropical Meteorology*, Miami, FL, Amer. Meteor. Soc., 240–241.
- Evan, A., and S. Camargo, 2011: A climatology of Arabian Sea cyclonic storms. *J. Climate*, **24**, 140–158.
- Evans, J. L., and M. P. Guishard, 2009: Atlantic subtropical storms. Part I: Diagnostic criteria and composite analysis. *Mon. Wea. Rev.*, **137**, 2065–2080.
- Fu, R., A. D. D. Genio, and W. B. Rossow, 1990: Behavior of deep convective clouds in the tropical Pacific deduced from ISCCP radiances. *J. Climate*, **3**, 1129–1152.
- , —, and —, 1994: Influence of ocean surface conditions on atmospheric vertical thermodynamic structure and deep convection. *J. Climate*, **7**, 1092–1108.

- Gadgil, S., V. Joseph, and N. V. Joshi, 1984: Ocean–atmosphere coupling over monsoon regions. *Nature*, **312**, 141–143.
- Graham, N. E., and T. P. Barnett, 1987: Sea surface temperature, surface wind divergence, and convection over tropical oceans. *Science*, **238**, 657–659.
- Gray, W. M., 1968: Global view of the origin of tropical disturbances and storms. *Mon. Wea. Rev.*, **96**, 669–700.
- , 1979: Hurricanes: Their formation, structure and likely role in the tropical circulation. *Meteorology over the Tropical Oceans*, Royal Meteorological Society, 155–218.
- 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 Reanalysis Project. *Bull. Amer. Meteor. Soc.*, **77**, 437–471.
- Kistler, R., and Coauthors, 2001: The NCEP–NCAR 50-Year Reanalysis: Monthly means CD-ROM and documentation. *Bull. Amer. Meteor. Soc.*, **82**, 247–267.
- 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). *Bull. Amer. Meteor. Soc.*, **91**, 363–376.
- Knutson, T. R., 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.
- 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. *Climate Dyn.*, **32**, 125–141.
- 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 intensities 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*, **23**, 2699–2721.
- Nolan, D. S., E. D. Rappin, and K. A. Emanuel, 2007: Tropical cyclogenesis sensitivity to environmental parameters in radiative-convective equilibrium. *Quart. J. Roy. Meteor. 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 20-km-mesh global atmospheric model: Frequency and wind intensity analyses. *J. Meteor. Soc. Japan*, **84**, 259–276.
- Ramsay, H. A., and A. H. Sobel, 2011: Effects of relative and absolute sea surface temperature on tropical cyclone potential intensity using a single-column model. *J. Climate*, **24**, 183–193.
- 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. *Atmos. Res.*, **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.
- Su, H., D. Neelin, and J. E. Meyerson, 2003: Sensitivity of tropical tropospheric temperature to sea surface temperature forcing. *J. Climate*, **16**, 1283–1301.
- Swanson, K. L., 2008: Nonlocality of Atlantic tropical cyclone intensity. *Geochim. Geophys. Geosyst.*, **9**, Q04V01, doi:10.1029/2007GC001844.
- Uppala, S. M., and Coauthors, 2005: The ERA-40 Re-Analysis. *Quart. J. Roy. Meteor. Soc.*, **131**, 2961–3012.
- Vecchi, G. A., and B. J. Soden, 2007a: Effect of remote sea surface temperature change on tropical cyclone potential intensity. *Nature*, **450**, 1066–1070.
- , and —, 2007b: Increased tropical Atlantic wind shear in model projections of global warming. *Geophys. Res. Lett.*, **34**, L08702, doi:10.1029/2006GL028905.
- , K. L. Swanson, and B. J. Soden, 2008: Whither hurricane activity? *Science*, **322**, 687–689.
- Villari, 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.*, **138**, 2681–2705.
- 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.
- , —, and J. C. L. Chan, 2009: Tropical cyclone genesis frequency over the western North Pacific simulated in medium-resolution coupled general circulation models. *Climate Dyn.*, **33**, 665–683.
- Yoshimura, J., M. Sugi, and A. Noda, 2006: Influence of greenhouse warming on tropical cyclone frequency. *J. Meteor. Soc. Japan*, **86**, 405–428.
- Zhang, C., 1993: Large-scale variability of atmospheric deep convection in relation to sea surface temperature in the tropics. *J. Climate*, **6**, 1898–1913.
- 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.