Detecting the ITCZ in Instantaneous Satellite Data using Spatiotemporal Statistical Modeling: ITCZ Climatology in the East Pacific

CAROLINE L. BAIN, JORGE DE PAZ, JASON KRAMER, GUDRUN MAGNUSDOTTIR, PADHRAIC SMYTH, HAL STERN, AND CHIA-CHI WANG*

University of California, Irvine, Irvine, California

(Manuscript received 9 March 2010, in final form 27 July 2010)

ABSTRACT

A Markov random field (MRF) statistical model is introduced, developed, and validated for detecting the east Pacific intertropical convergence zone in instantaneous satellite data from May through October. The MRF statistical model uses satellite data at a given location as well as information from its neighboring points (in time and space) to decide whether the given point is classified as ITCZ or non-ITCZ. Two different labels of ITCZ occurrence are produced. IR-only labels result from running the model with 3-hourly infrared data available for a 30-yr period, 1980–2009. All-data labels result from running the model with additional satellite data (visible and total precipitable water), available from 1995 to 2008. IR-only labels detect less area of ITCZ than all-data labels, especially where the ITCZ is shallower. Yet, qualitatively, the results for the two sets of labels are similar.

The seasonal distribution of the ITCZ through the summer half year is presented, showing typical location and extent. The ITCZ is mostly confined to the eastern Pacific in May, and becomes more zonally distributed toward September and October each year. Northward and westward shifts in the location of the ITCZ occur in line with the seasonal cycle and warm sea surface temperatures. The ITCZ is quite variable on interannual time scales and highly correlated with ENSO variability. When the ENSO signal was removed from labels, interannual variability remained high. The resulting IR-only labels, representing the longer time series, showed no evidence of a trend in location nor evidence of a trend in area for the 30-yr period.

1. Introduction

The intertropical convergence zone (ITCZ) is one of the most recognizable aspects of the global circulation that can be seen from space. The ITCZ forms as a zonally elongated band of cloud at low latitudes where the northeasterly and southeasterly trade winds converge. The focus of this paper is to introduce a new method that detects the ITCZ using high temporal and spatial resolution satellite products that have recently been archived. We apply the method in the Pacific region north of the equator and east of the date line during the boreal summer half year (May through October, hereafter summer).

Previous observational studies of the global climatological ITCZ (e.g., Mitchell and Wallace 1992; Waliser and Gautier 1993) focused on the annual cycle in different regions. They found very distinct longitudinal variations in the ITCZ. In the Indo–western Pacific region the summer ITCZ is broad in latitude and ill-defined due to the extensive warm pool in the ocean and monsoonal circulations. However, in the east Pacific the mean summer ITCZ is narrow and long, generally located at the southern boundary of the east Pacific warm pool, north of the strongest meridional gradient of sea surface temperature (Raymond et al. 2003). During the summer the east Pacific ITCZ is particularly visible in instantaneous satellite fields. During Northern Hemisphere winter the ITCZ remains in the Northern Hemisphere, but its signature is considerably weaker and gets mixed in with signatures of extratropical frontal systems owing to cold air outbreaks (Wang and Magnusdottir 2006, hereafter WM06).

Traditionally, the ITCZ has been identified in terms of time-averaged fields, either in terms of the seasonal mean outgoing longwave radiation (OLR) or, in more recent

* Current affiliation: Department of Atmospheric Sciences, Chinese Culture University, Taipei, Taiwan.

Corresponding author address: Gudrun Magnusdottir, 3321 Croul Hall, Dept. of Earth System Science, University of California, Irvine, Irvine, CA 92697.
E-mail: gudrun@uci.edu

DOI: 10.1175/2010JCLI3716.1

years, in terms of the seasonal mean precipitation. For example, Waliser and Gautier (1993) used thresholding of mean OLR in combination with mean high reflectivity to identify the ITCZ. A low OLR threshold is chosen, and values below this threshold are considered to represent clouds that are part of the convection within the ITCZ. To filter out smaller systems an area limit is often imposed. To date, little attempt has been made to group the low OLR clouds into a continuous zonally elongated feature nor to remove noise and accommodate shallower convection, which often characterizes the narrow east Pacific ITCZ region (Zhang et al. 2004).

Other studies (e.g., Serra and Houze 2002), have referred to the ITCZ as a geographical region along which westward propagating disturbances (WPDs) tend to propagate zonally. Weak WPDs (or easterly waves) have been observed to propagate through the ITCZ cloud envelope (e.g., Scharenbroich et al. 2010) and, in general, it is impractical to attempt to separate the easterly wave signal out of the ITCZ signal. Magnusdottir and Wang (2008) reached the same conclusion when using spectral analysis on 40-yr European Centre for Medium-Range Weather Forecasts Re-Analysis (ERA-40) 850-hPa relative vorticity. The line of thinking that easterly waves or WPDs are inseparable from the ITCZ accommodates the idea that the ITCZ is composed of WPDs, but Gu and Zhang (2002a,b) demonstrated there is an additional important component to the ITCZ that is not propagating.

Satellite images show that from day to day the ITCZ is highly dynamic and changeable. The ITCZ can form as a narrow band of convection stretched over an extensive longitudinal distance (up to 70°) for hours to days, until its structure breaks down into individual disturbances that may move away or dissipate in place. The ITCZ envelope of convection usually re-forms within a day or two of breakdown (WM06). This dynamic ITCZ has been the subject of dynamical modeling studies (Ferreira and Schubert 1997; Wang and Magnusdottir 2005; Wang et al. 2010) that have concluded that barotropic instability is important in ITCZ breakdown.

The variability of the ITCZ presents a serious challenge to its automatic detection in instantaneous data. Here we want to focus on the ITCZ as a weather feature that has long been recognized by satellite meteorologists who analyze instantaneous fields. We use the following criteria to define the ITCZ.

(i) The ITCZ is a predominantly zonal feature.
(ii) It is cloudy but there may be cloud-free regions within the envelope of convection and the convection may be shallow (as represented by rather warm cloud-top temperatures).
(iii) The ITCZ is a large-scale feature and isolated tropical disturbances, unconnected to larger cloudy regions, are not part of the ITCZ.

We use satellite fields of infrared (IR), visible (VS), and total precipitable water (TPW) to find large-scale zonally connected regions of convection.

WM06 present a detailed study of daily variability of the ITCZ in the east Pacific for five years, 1999–2003, using visual inspection of satellite images. Here we extend their study by presenting a new method to automate the detection so that a greater number of active seasons may be analyzed. The new method is based on a statistical model that will produce the same results given the same data, which may not be true for human identification of the ITCZ. In other respects, the model seeks to emulate human analysis of the meteorological feature—looking for continuity in space and time while accommodating different meteorological fields. The output from the statistical model is a binary field, referred to as ITCZ labels, representing the presence (by 1) or absence (by 0) of the ITCZ in the area of interest at each time point. In this paper we introduce and validate the statistical model as well as present first results focusing mainly on the climatology of the ITCZ, its interannual variability and seasonal evolution.

The paper is organized as follows: Section 2 describes the input satellite data, section 3 the statistical model, and section 4 evaluates our approach to ITCZ detection. Section 5 presents results from two different outputs from the model: a longer (1980–2009) time series of ITCZ labels obtained using IR data only and a shorter (1995–2008) time series of labels obtained using all three sources of data (IR, VS, and TPW). Section 6 presents concluding remarks.

2. Satellite data

Satellite IR, VS, and TPW images from 90°W to 180° were used for analysis. The VS data are a measure of reflectivity of clouds, the IR data correspond to cloud-top temperature. The TPW data give the water vapor content of the entire column in units of equivalent thickness of liquid water (0–75 mm).

The IR and VS datasets were obtained from the National Climatic Data Center (NCDC) of the National Oceanic and Atmospheric Administration. These two datasets are from the hurricane satellite (HURSAT) Basin data (Knapp 2008) and are closely related to HURSAT B1 data (Knapp and Kossin 2007), which are derived from International Satellite Cloud Climatology Project (ISCCP) B1 data. The data were collected from radiometers on different geostationary satellites. The IR channel data were

recalibrated to reduce intersatellite differences. The HURSAT Basin data have the same spatial and temporal resolution as HURSAT B1 data, but are assembled, gridded, and archived for each tropical ocean basin separately (more information available online at http://www.ncdc.noaa.gov/oa/rsad/hursat/). The IR dataset is available every three hours (0000 UTC, etc.) from 1980 to 2009 at a spatial resolution of 10 km. We chose to reduce the spatial resolution of the satellite fields to 0.5° in order to reduce the number of grid points and decrease computation time. Coarsening the grid does not influence the end result as the ITCZ objects we identify are considerably larger.

The longitudinal extent of our domain (90°W–180°) means that the entire area is in daylight only once per day (2100 UTC) even though global VS data are available every three hours. Thus the VS dataset is only useful once per day; however the statistical model is able to accommodate for the missing measurements. We also found that the earlier parts of the record were problematic for the VS channel; therefore, the VS dataset is only used from 1995 to 2008.

The TPW data was assembled and archived by Remote Sensing Systems (more information available online at www.remss.com). The data is a composite of available microwave data [from the Special Sensor Microwave Imager (SSM/I), Tropical Rainfall Measuring Mission (TRMM) Microwave Imager (TMI), Advanced Microwave Scanning Radiometer for Earth Observing System (EOS; AMSR-E) satellites]. It covers 1987–2008; however, the earlier part of the record is at lower temporal resolution so we only use the data for 1995–2008. The data is at a spatial resolution of 0.25° that we coarsen to 0.5°. The images are from 0000, 0600, 1200, and 1800 UTC and contain data from ±3 h, in practice making them ±9 h resolution at each recorded time point. Furthermore, we applied linear interpolation between the time points, making TPW available at the same time as IR.

To summarize, one dataset, IR, is available from 1980 to 2009. The other two datasets, VS and TPW, are used for the latter part of the time period, 1995–2008. We therefore chose to run our analysis of the east Pacific ITCZ twice. First we used all three satellite datasets for the period 1995–2008. Second, we ran our analysis for the entire time period, 1980–2009, using the IR data only.

Since there are two versions of the model with different inputs, it is critical that we have consistent terminology to identify the output. The term ''all-data labels'' refers to the set of ITCZ labels obtained from the Markov random field (MRF) model using IR, VS, and TPW satellite data as input. The term ''IR-only labels'' refers to the set of ITCZ labels obtained from the MRF model using only IR for input.

3. Method description

The statistical model used to detect the ITCZ incorporates satellite data, a Markov random field that represents ITCZ/non-ITCZ, and a priori information on the most likely location of the ITCZ. In the following explanation we will use X_{ijt} as an indicator of the presence ($X_{ijt} = 1$) or absence ($X_{ijt} = 0$) of the ITCZ at longitude i, latitude j, and discretized time t. The corresponding satellite data will be denoted Y_{ijt}, a vector comprised of satellite IR, VS, and TPW. The indicator X_{ijt} is not directly observable since the ITCZ is not defined deterministically as a function of satellite data (such as clouds below some temperature are always ITCZ, etc.). Thus, the binary status of ITCZ at each grid point is instead inferred from satellite data, Y_{ijt}, and the status of surrounding grid points in time and space. The final ITCZ status of each grid point is determined by sampling from the posterior distribution:

$$P(X|Y) = \frac{P(Y|X)P(X)}{\sum_X P(Y|X)P(X)}, \tag{1}$$

where $P(Y|X)$ gives the likelihood of the satellite data Y, given X (ITCZ or non-ITCZ status). Then $P(X)$ is a prior distribution on the probability of ITCZ that incorporates neighborhood and spatial information. We describe the specifics of this model, our computational approach, and other implementation details in the remainder of this section.

a. Learning the distribution of satellite data, $P(Y|X)$

The model requires knowledge of the satellite values that are characteristic of ITCZ and non-ITCZ. This information was obtained from training data provided by manual labeling of satellite data. Three meteorologists identified closed ITCZ regions (if any) at 2100 UTC for each day during August 2000. The labelers were provided with a graphical user interface and a tablet board with stylus to maximize precision when labeling the region desired. They were given VS, IR and TPW at times $t - 2$, $t - 1$, t, $t + 1$ and $t + 2$ where t was the timeframe being labeled. The individuals using the tablet board were told to label based on principles listed in the introduction (section 1). Continuity in time and reconciling the different meteorological fields are important elements in the manual labeling and constitute what the statistical model seeks to emulate.

The mean vector and covariance matrix for VS, IR, and TPW, for grid points belonging to the ITCZ ($X = 1$) and non-ITCZ ($X = 0$) were created using the manual labeled data. The means and covariances were used to produce two Gaussian distributions for the satellite data

TABLE 1. Table showing the comparisons between a union of the manual labels defined as ITCZ and each method of detection for August 2000. False negative is the number of ITCZ grid points not detected by each method divided by the number of ITCZ grid points. False positive is the number of grid points incorrectly detected as ITCZ by each method divided by the number of ITCZ grid points in the union. Results given as percentages where low numbers indicate better agreement between methods, except for the final column where a higher number indicates better agreement. See text for more details.

Method tested	Base method tested against	False negative	False positive	Total absolute error	Intersection/union
Person P1	Union P2, P3	5.9	2.3	8.2	57.9
Person P2	Union P1, P3	7.2	2.8	10.1	51.4
Person P3	Union P1, P2	7.7	2.2	9.9	49.3
All-data labels	Union P1, P2, P3	7.5	4.4	11.9	49.2
IR-only labels		11.3	2.6	13.9	37.9
Threshold IR		12.3	3.7	16.0	32.6
Threshold TPW		6.7	9.8	16.4	43.4
If all domain "ITCZ"		0.0	78.8	78.8	21.2
If all domain "non-ITCZ"		21.2	0.0	21.2	0.0
All-data labels	Threshold IR	1.9	7.3	9.2	53.5
IR-only labels		2.2	2.0	4.2	67.5
Threshold TPW		2.7	14.3	17.0	35.4
If all domain ITCZ		0.0	87.3	87.3	12.7
If all domain non-ITCZ		12.7	0.0	12.7	0.0
All-data labels	Threshold TPW	8.2	1.9	10.0	57.3
IR-only labels		13.9	2.0	15.9	36.6
Threshold IR		14.3	2.7	17.0	35.4
If all domain ITCZ		0	75.6	75.6	24.4
If all domain non-ITCZ		24.4	0	24.4	0

Y, conditioned on $X = 1$ and $X = 0$. These distributions (for each VS, IR, and TPW) had different means and variances for ITCZ ($X = 1$) and non-ITCZ ($X = 0$), but also overlapped. If the satellite data at a given grid point falls within the overlap regions, there is less certainty as to whether $X = 1$ or $X = 0$. The Gaussian distribution for $X = 1$ is less broad than the one for $X = 0$, as the ITCZ is mostly cloudy, whereas the non-ITCZ is a mixture of cloud and noncloud. It is assumed that, given X, the observed satellite values, Y, are independent across different grid points and times, so that

$$P(Y|X) = \prod_{ijt} P(Y_{ijt}|X_{ijt}). \quad (2)$$

The set of manual labels that performed best against the union of labels (Person 1, see Table 1) was used to train the statistical model. Additional months were subsequently labeled to check that $P(Y|X)$ did not change significantly, but these labels were not used for training the final model.

b. A Markov random field model for the ITCZ, P(X)

The Markov random field is a mathematical model (Kindermann and Snell 1980; Geman and Geman 1984; Li 1995; Smyth 1997) that specifies a probability distribution on the spatial characteristics of the ITCZ. The MRF allows one to incorporate typical spatial locations for the ITCZ along with the informal notion that, if one grid point is part of the ITCZ, then this ought to increase the likelihood that neighboring grid points are also part of the ITCZ.

In an MRF model the status of a single grid point, X_{ijt}, depends only on grid points that are immediate neighbors. If we define N_{ijt} as the set of neighbors of X_{ijt}, and X_{-ijt} as the set of all grid points except our initial grid point, then the Markov property says that

$$P(X_{ijt}|X_{-ijt}) = P(X_{ijt}|N_{ijt}). \quad (3)$$

The neighborhood that we use includes the immediate neighbors in the longitude ($X_{i+1,j,t}$, $X_{i-1,j,t}$), latitude ($X_{i,j+1,t}$, $X_{i,j-1,t}$), and time ($X_{i,j,t+1}$, $X_{i,j,t-1}$) directions. Figure 1 is a schematic of the neighborhood structure. The MRF can also incorporate a bias or tendency for particular locations to be part of the ITCZ. This is useful because we know the ITCZ to be a zonal feature that occurs in tropical latitudes only. To implement this we specify a spatial prior, q_j, which does not vary in time. We used a longitudinally constant spatial prior with Gaussian dependence on latitude. Figure 2a shows the study region. Figure 2b shows the two spatial priors used for the different ITCZ labels. The spatial prior has its highest probability density for latitudes between 7.5° and 10°N. The values decay as a Gaussian curve, with a variance of 3°, to a minimum value, $q_j = 0.05$. This is nonzero to allow the model to diagnose ITCZ in unlikely regions if the signal is particularly strong. Including TPW as an input tends to link up the convective regions; therefore higher probability

values were required for the spatial prior used to create the IR-only labels. These values were obtained through a series of sensitivity studies during model validation.

Incorporating all of this information, the conditional probability that characterizes our MRF can be written as

$$P(X_{ijt}=1|X_{-ijt}) \propto \exp[\beta_i I(X_{i-1,j,t}=1) + \beta_i I(X_{i+1,j,t}=1) + \beta_j I(X_{i,j-1,t}=1) + \beta_j I(X_{i,j+1,t}=1) + \beta_t I(X_{i,j,t-1}=1) + \beta_t I(X_{i,j,t+1}=1) + \beta_s \log q_j], \quad (4)$$

where I is an indicator function equal to 1 if the condition inside the parentheses is true, and where β_i, β_j, β_t, and β_s determine the strength of relationship between each grid point and its zonal, meridional, and time neighbors and its spatial location, respectively. The probability is written as proportional to the specified term because calculating the actual probability requires computing a similar expression for the probability that $X_{ijt} = 0$ and then normalizing so that they sum to one. Based on some experimentation with a sample of training data we set all β values equal to 1. The analysis results are not sensitive to β values. Rather than optimizing and risk overfitting to the training data, we opted to select a simple intuitive set of values. Interestingly, strengthening the β in the zonal direction did not necessarily improve the ITCZ output and results obtained by altering the strength of the time or spatial β did not provide systematic improvements over using $\beta = 1$.

c. *Inferring the presence/absence of ITCZ, P(X|Y)*

We take a Bayesian approach to obtain our data-driven inferences of the ITCZ. The probability model based on training data, given ITCZ status $P(Y|X)$, is combined with the MRF prior distribution on ITCZ status, $P(X)$, to obtain the posterior distribution for the ITCZ status $P(X|Y)$ (Gelman et al. 2003). A widely used technique for making inferences about X given Y, the spatial prior and the model parameters, is via a Markov chain Monte Carlo (MCMC) algorithm known as Gibbs sampling (Geman and Geman 1984; Gilks et al. 1993). Initial values are first generated for X (each X_{ijt} is set to either 0 or 1) based on the satellite data. This is done by stochastically assigning X_{ijt} to be 1 with probability

$$\frac{P(Y_{ijt}|X_{ijt}=1)}{P(Y_{ijt}|X_{ijt}=1)+P(Y_{ijt}|X_{ijt}=0)} \quad (5)$$

and 0 otherwise (where we have assumed that X_{ijt} is a priori equally likely to be zero or one for purposes of this initialization). This is an initial approximation based on the satellite data alone. These initial values are then updated sequentially with each grid point updated conditionally on the values of its neighbors. This is done via a conditional probability calculation that is analogous to the MRF probability given in Eq. (4) except that it now includes a term for the satellite data,

$$P(X_{ijt}|X_{-ijt}, Y_{ijt}) \propto P(Y_{ijt}|X_{ijt}) \times P(X_{ijt}|X_{-ijt}), \quad (6)$$

where the second term on the right is the rhs side of Eq. (4). Gibbs sampling requires cycling through all grid points to update their values repeatedly, using the most recent values for all neighboring grid points. Iterations through all grid points are repeated until the resulting Markov chain is found to have converged to stationary behavior. This stationary behavior reflects the desired distribution $P(X|Y)$. We found empirically that 200 iterations were sufficient to obtain stationary behavior. The next 50 simulations are then taken as representatives of the posterior distribution. Grid points that are

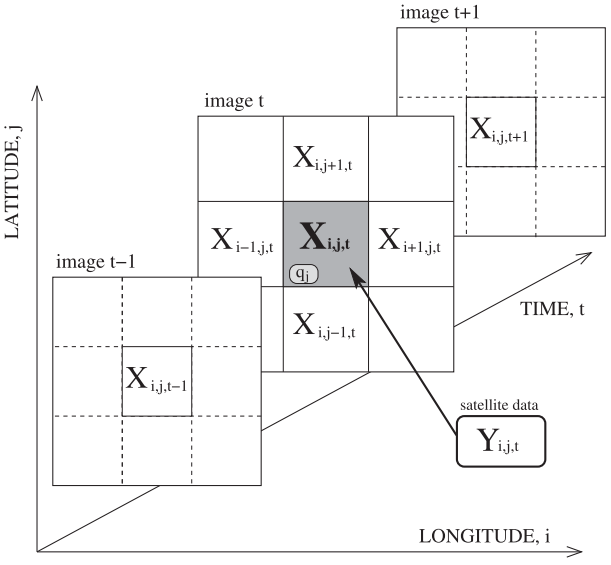

FIG. 1. Schematic of the neighborhood structure used in the MRF: X is the status of the ITCZ or non-ITCZ at grid point i, j, t; Y is the satellite data; and q is the spatial prior value at the grid point. The binary status of $X_{i,j,t}$ depends on $Y_{i,j,t}$, the status of X in the neighboring grid points in space and time, and q_j.

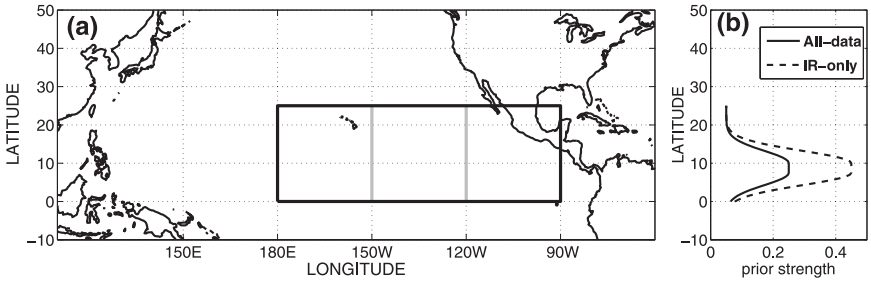

FIG. 2. (a) Map of the study region (black rectangle) split into three longitudinal boxes in gray (used in section 5). (b) Spatial prior for the model for All-data labels (solid) and IR-only labels (dashed) as a function of latitude. The spatial prior gives a starting point for the probability that gridpoint locations are part of the ITCZ. Each spatial prior is zonally uniform across the domain.

part of the ITCZ in 50% or more of these 50 simulations are defined as ITCZ in the final output. The simulations also provide information about the probability of ITCZ being present at each grid point.

The calculation described here is computationally intensive. To improve efficiency, our model recognizes the neighborhood structure as a three-dimensional chess board with each grid point as a square and time serving as the third dimension. The Gibbs sampling update for the ''black'' squares depends only on the status of neighboring ''white'' squares and vice versa. This means that we can use a vector calculation to simultaneously update all of the ''black'' squares in a single step rather than use grid point–by–grid point updating. Parallelizing the computations of disparate time points is also feasible, although not explored in this paper.

d. Postprocessing

Postprocessing is performed on the model output as a form of quality control. Since we are looking for a large continuous feature, the postprocessing was designed to eliminate any noise not removed by the MRF. Postprocessing on the ITCZ labels is applied in three steps: 1) ITCZ regions are closed by dilation and erosion (documentation on erosion and dilation algorithms at www.mathworks.com/help/toolbox/images/ref/bwmorph.htm), joining disjointed areas and smoothing edges; 2) any holes are filled in as it is assumed that, if a region is surrounded by ''ITCZ,'' then it also must be part of it; and 3) Any identified ''on'' regions smaller than 100 grid points (approx 5° by 5°) are removed.

4. Evaluation of ITCZ detection

Figure 3 shows an example of the ITCZ as defined by manual labels, the statistical model and IR thresholding. The figure demonstrates that both the manual label and the statistical model incorporated regions of variable cloud-top temperatures, grouping together the envelope of the convective region. The thresholding only uses clouds below a certain temperature as ITCZ. Since the ITCZ has not been definitively identified as a binary signal in instantaneous data, it is difficult to assess the validity of the statistical model. Thus, we have chosen to evaluate the model against the other techniques available to us—manual labels and thresholding of IR and TPW. Different values of thresholds were tested until an optimum value (which minimized errors in comparison to the union of manual labelers) was found. For IR this was a cloud-top temperature threshold of 270 K and for TPW a value of 53 mm. Postprocessing was performed on the thresholded returns such that only cloud systems exceeding 200 km in perimeter were used (approximately similar to Garcia 1985).

Figure 4 describes how the labels were evaluated. A base method considered to be ground truth and a test method were defined. The test method was scored against the base method by identifying how many grid points were wrongly identified. Referring to the diagram,

$$\text{false negative} = C/(A + B + C + D), \tag{7}$$

$$\text{false positive} = B/(A + B + C + D), \tag{8}$$

and

$$\text{total absolute error} = (B + C)/(A + B + C + D), \tag{9}$$

where A is the region where both the base and test methods agree there is ITCZ, B is ITCZ in the test method but not the base method, C is ITCZ in the base method and not in the test method, and D is defined as non-ITCZ by both methods. If the base method and the test method are similar, then all three scores are low. In addition to these error measures, a measure of agreement of overlap that is not affected by the size of the domain is

$$\text{intersection/union} = A/(A + B + C). \tag{10}$$

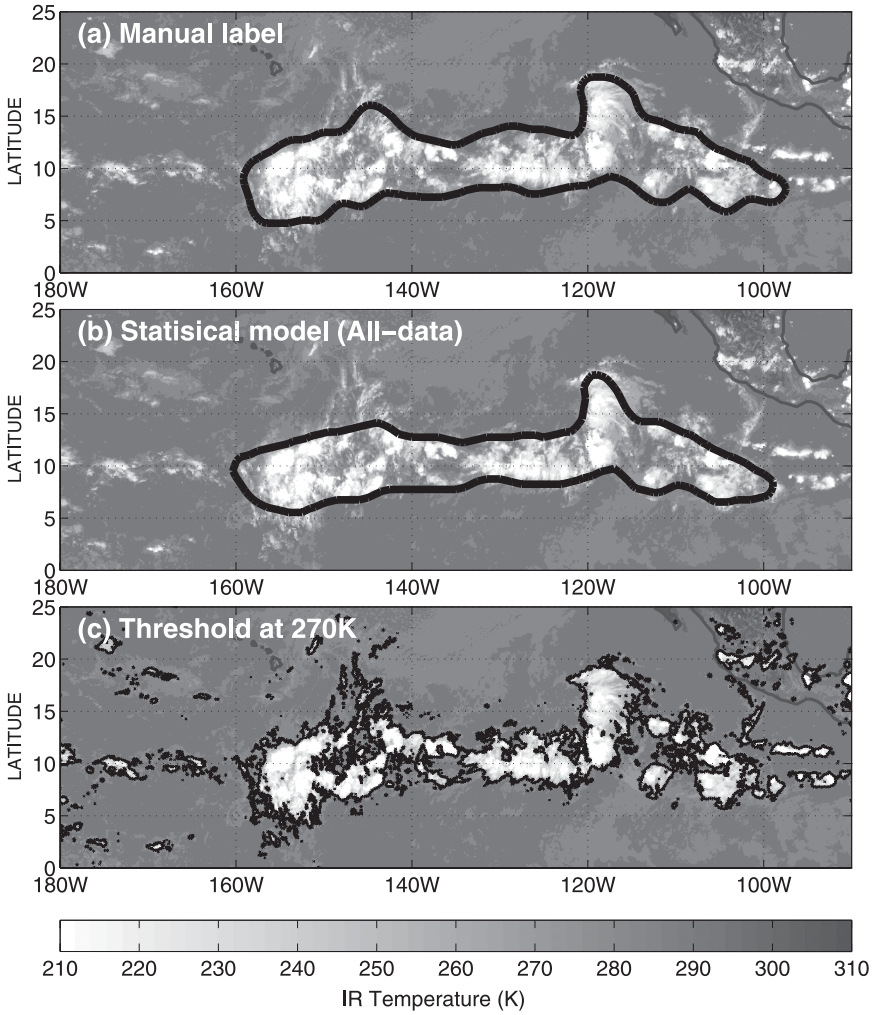

FIG. 3. ITCZ label for 2100 UTC 19 Aug 2000 overlaid on the IR image: (a) manual label (union of three labelers), (b) statistical model using IR, VS, and TPW satellite data, and (c) threshold at 270 K (without postprocessing). Shading represents the IR image with the temperature scale provided below. The bold contours mark the ITCZ labels.

This measure is 100 for perfect agreement and 0 when no grid points are identified by both methods as ITCZ (no overlap).

Scores were calculated for each image in the evaluation dataset, and Table 1 shows the mean of these errors over the whole month of August 2000. The first block is a comparison of each manual labeler against a union of the other labelers. For example Person 1 labels are compared to the union of labels from Person 2 and Person 3. The second block is a comparison between automated methods and the union of labels from persons 1, 2 and 3. The final blocks of the table compare the various automated methods to each other. The results are obtained by comparing labels in the region from 0° to 20°N, 170° to 100°W. It was found that the manual labelers were unlikely to label to the edges of the full domain as they could not see past them, so a narrow comparison region was a more reasonable choice.

One limitation of the error rates is that they can be hard to calibrate: a low score is good but it is not obvious how

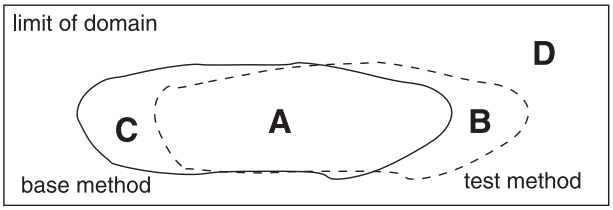

FIG. 4. Schematic showing the areas used to assess the agreement between ITCZ labels found using different methods. The outline of the ITCZ is shown by the solid curve for the base method and by the dashed curve for the test method.

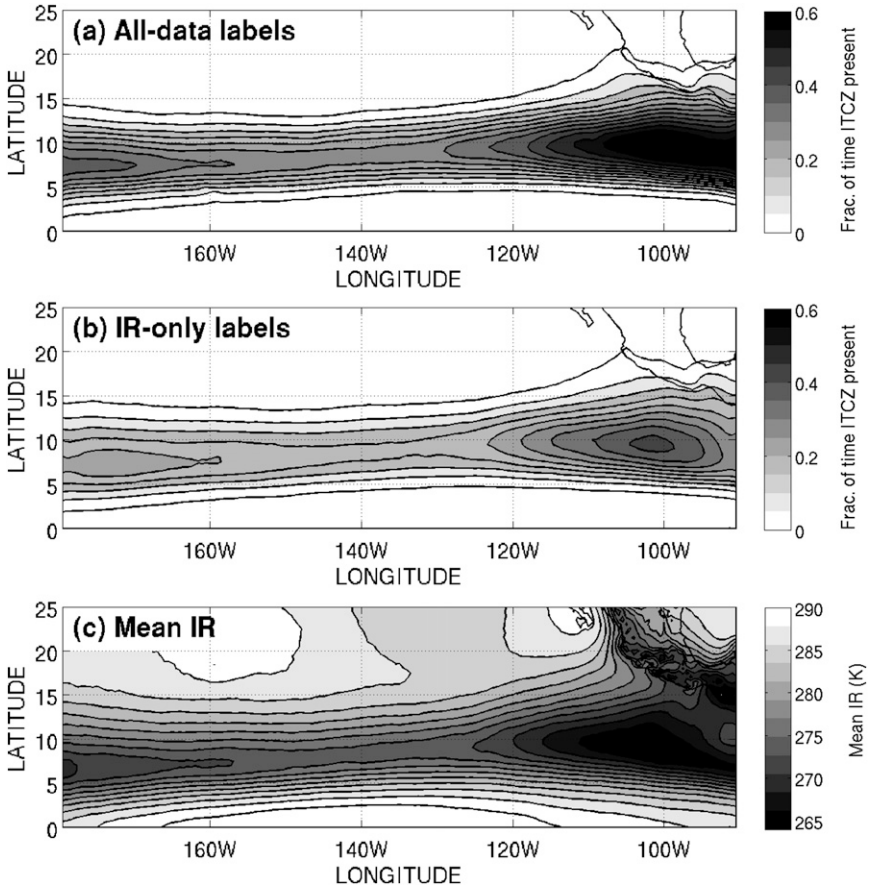

FIG. 5. Fraction of time when the ITCZ is present during May–October from 1995–2008 with (a) all-data labels and (b) IR-only labels. (c) Mean IR is shown for comparison: cold temperatures indicate higher cloud tops.

to recognize an unacceptably high score. For example, a total error score of 100 could only be obtained if the test image were the negative image of the base method image, which is not at all likely. To put the scores in perspective, the error scores for "all domain defined as ITCZ" and "all domain defined as non-ITCZ" are also given. The first three columns are the error scores, where a lower score indicates more skill for the method tested. The results show that, overall, all-data labels are more similar to manual labels on all measures than either thresholding method. Thresholded IR generally underestimates the ITCZ region and thresholded TPW generally overestimates. The IR-only labels somewhat underestimate the ITCZ region, but not by as much as thresholding IR. The intersection/union score is shown in the final column of Table 1, where a higher score indicates more skill. Again, the all-data labels score best in comparison to the other automated methods. The IR-only labels are an improvement on thresholded IR, but do not score as well as thresholded TPW. The bottom section of Table 1 compares the various automatic methods with each other. The labels generated by the MRF model are more similar to each than the thresholding techniques are to each other. The model labels seem to capture much of the information in the thresholding approaches (they seem to agree fairly often) and as described above do better at matching the human labelers.

Therefore, by these evaluations, the ITCZ labels generated by the statistical model perform better than the other automated techniques against the manual labels by experts, as well as against the other automated techniques. This justifies using the statistical model output to assist in studying the climatology and dynamics of the ITCZ.

5. Results: Climatology of the ITCZ

The labels of ITCZ show much interannual, intraseasonal, and synoptic-scale variability, emphasizing the complex dynamic nature of the weather feature. In this section we present the composite picture of the ITCZ from all-data labels (available 1995–2008) to provide the basic climatology and interannual variability of ITCZ.

We use IR-only labels (available 1980–2009), to investigate long-term climatic trends.

Figure 5 shows the mean location of the ITCZ obtained from the sets of ITCZ labels for the same time period (1995–2008) alongside mean IR for the same period. The plot shows that qualitatively the labels show similar locations but they are quantitatively different. The locations are also comparable to the location of cold mean IR. The ITCZ is diagnosed less when only IR is used as input. The correlation between the IR-only and all-data labels is 0.989. The all-data labels produce a field that is on average 1.3 times the frequency obtained from the IR-only labels in the region 0°−15°N (the agreement is higher in the northern portion where there was no ITCZ present). The IR-only labels give a more conservative estimate of the ITCZ because VS is not included, which may discount low cloud. In addition TPW, which is a smooth field linking up convective regions is not included. Despite this, Fig. 5 demonstrates that the mean location is reproduced for both sets of labels, giving confidence that statements made about general characteristics from averaged fields are consistent across the two label datasets.

a. Seasonal evolution of the ITCZ

The time evolution of the ITCZ through the season from the all-data labels for 1995–2008 is shown in Fig. 6. The shading represents mean width (degrees latitude) of the ITCZ at each longitude. The figure indicates that the ITCZ is most present east of 130°W at the beginning of the season, growing steadily in width, reaching an initial peak in late June. As the boreal summer months progress, the ITCZ becomes more zonally distributed. The ITCZ west of 150°W becomes wider from the beginning of August until the end of October. In the eastern part of the domain there is a secondary peak in width in late August, and then a decline in size in September and October.

Figure 6b shows the time evolution of mean area of cloud associated with tropical cyclones for the whole domain, found using the algorithm described in the appendix. Tropical cyclone cloud peaks in August and September, coinciding with a recovery of ITCZ amount in the east Pacific (the double-peak effect). Therefore, it is likely that some of the ITCZ defined in the east Pacific at this time is attributable to tropical cyclone activity, which is mostly confined to these longitudes. Another reason for a recovery from the decline in the ITCZ in late August may be the increase in the number of WPDs associated with African easterly waves, which peak annually in August and September (Thorncroft and Hodges 2001).

This image was reproduced using the IR-only labels as well as the thresholded IR and thresholded TPW datasets (not shown), and that the features observed are consistent

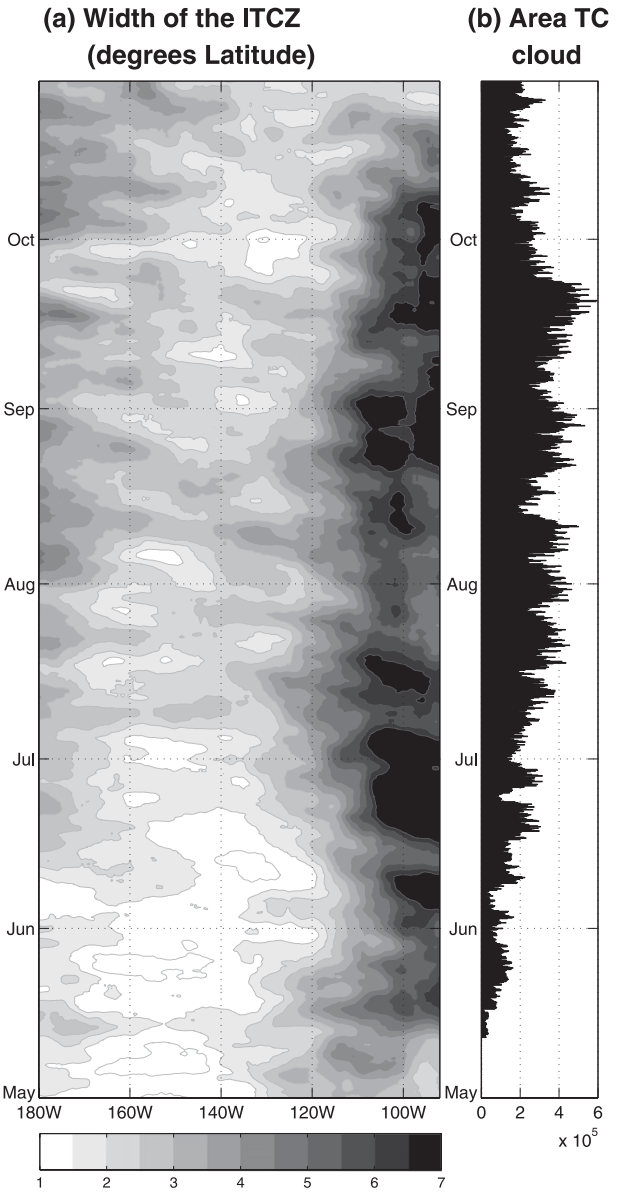

FIG. 6. (a) Mean width of ITCZ (°lat) vs time of year (shaded) using all-data labels (1995–2008). (b) Mean area of cloud (km^2) associated with tropical cyclones vs time of year from 1995 to 2008.

between datasets. The plots are qualitatively similar: the double-peak in the eastern Pacific ITCZ width is always detected, as is the westward spread of ITCZ activity toward the end of the season. The main differences between different labels are in the central region (120°–150°W) where clouds are lower and warmer and, therefore, less likely to be detected in labels that only rely on IR data.

Figure 7 shows seasonal-mean sea surface temperature (SST) and ITCZ fields followed by monthly anomalies of ITCZ location overlaid on anomalies of SST. The median anomaly was used for averaging over the years, rather

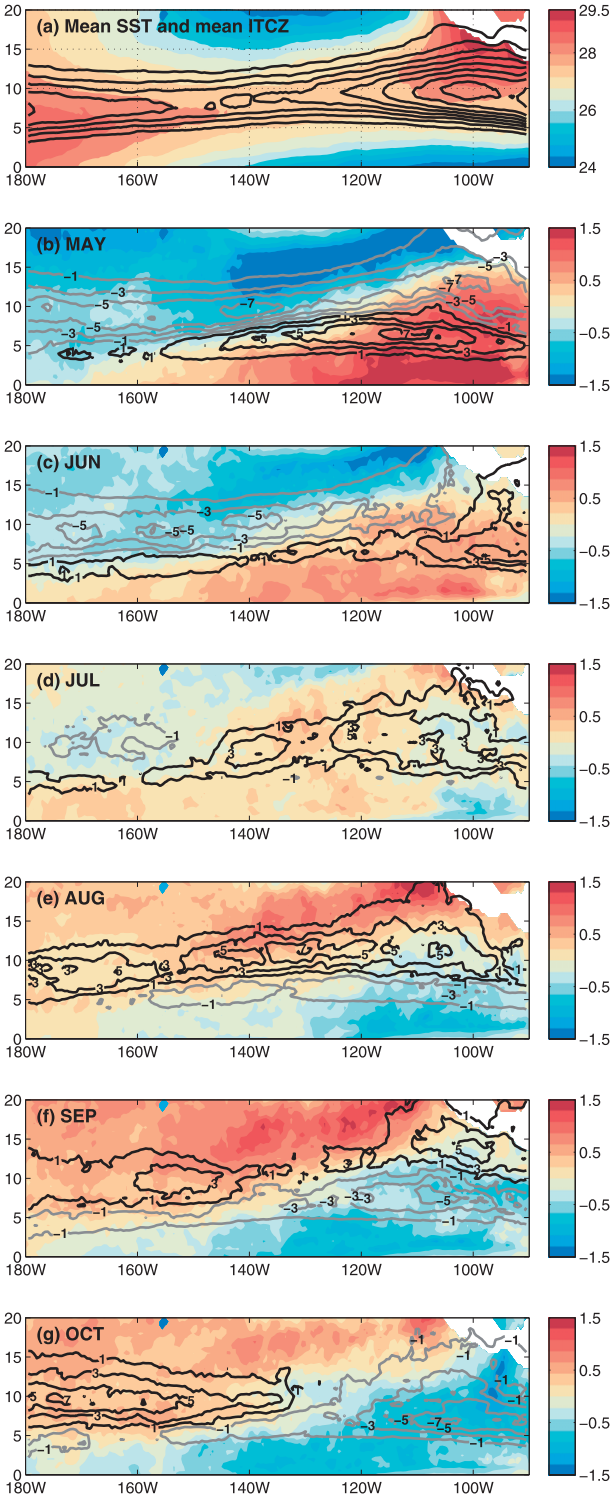

FIG. 7. (a) Mean SST (shaded) with mean ITCZ days per season contoured every 20 days. (b)–(g) Monthly anomalies of ITCZ location (contoured) overlaid on monthly anomalies of SST. All averaged from 1995 to 2008. The contours represent median number of days in the month when the ITCZ is present in comparison to the mean. Black (gray) curves are positive (negative) anomalies; contour interval for the anomalies is 2 days. SST (°C) is as shown in the color bars.

than the mean anomaly, to reduce the influence of the El Niño year in 1997 when the ITCZ was located particularly far south toward the end of the season.

The mean SST shows a sloping west-southwest to east-northeast orientation in warm temperatures, with higher temperatures in the far west and far eastern parts of the domain. Raymond et al. (2006) stated that ITCZ formation in the east Pacific was most likely to occur in the region of maximum SST gradient on the southern boundary of the SST maximum. On inspection of the SST data, this location often coincides with the location of the ITCZ in the region from the Central American coastline to 120°W. The relationship is the reverse in the central Pacific (approximately 140°W to edge of domain bounding box at 180°), where we typically find the ITCZ on the maximum SST gradient north of the SST maximum. In between these regions there is a transition zone from 120° to 140°W where the ITCZ coincides with the region of highest SSTs. It should be noted that this transition zone overlaps with the region, 130° to 160°W, where the clouds are shallower (warmer IR) and the ITCZ is least detected if total ITCZ occurrence for each longitude is calculated.

Figures 7b–g show the correspondence between anomalies in SST and the ITCZ. The figure demonstrates the migration of the ITCZ northward, reaching its maximum latitude in August and September. The redistribution of the ITCZ to the west can also be observed toward the end of the season. Positive anomalies can be seen in the eastern Pacific in June and July; these become negative in September and October. These seasonal shifts in location are consistent with the northward and westward migration of warm SSTs.

Figure 8 shows the median latitudinal distribution of the ITCZ for May to October from all-data labels. Again, the median was used to reduce the influence from the extreme 1997 El Niño year. The eastern and western longitudinal regions are shown to highlight the difference between the eastern and central Pacific. A tilt in the ITCZ from southwest to northeast is evident from the graphs. This tilt has been found in other observational studies (e.g., WM06; Magnusdottir and Wang 2008) and in idealized studies (Wang and Magnusdottir 2005; Ferreira and Schubert 1997), where it arises from advection by the forced vorticity field. It is also shown that the east Pacific ITCZ covers a larger region at the beginning of the season, but the envelopes of standard deviation around both curves demonstrate the considerable variability in the area of the ITCZ in May. There is a northward shift from May to September of approximately 3° latitude for both east and west domains, but the southward tilt to the west remains throughout the season. The redistribution toward the end of the season is also clear in this plot.

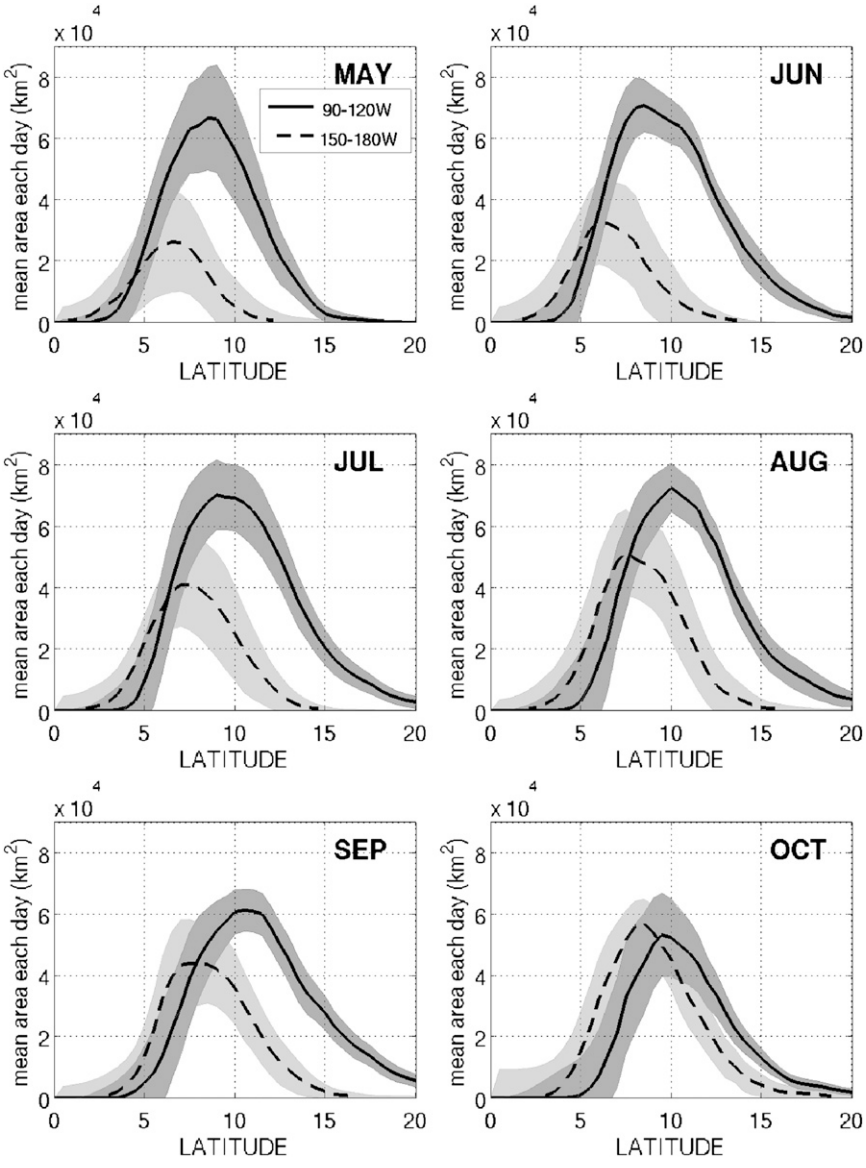

FIG. 8. Median latitude of ITCZ distribution in two longitudinal box regions, 90°–120°W (bold curve) and 150°W–180° (dashed curve): std dev shaded.

b. Interannual variability

The all-data labels were used to examine interannual variability in location, shown in Fig. 9. The plots are anomalies from the mean location of ITCZ (Fig. 5a). The figure shows considerable variability in the location of the ITCZ from year to year and especially demonstrates the strong influence of El Niño. Particularly noticeable are the El Niño events in 1997 and 2002 when the ITCZ region increased both in size and frequency of occurrence. The El Niño in 1997 was particularly notable for the large average area of ITCZ—this was also a strong El Niño event (McPhaden 1999). The La Niña events were in 1995 (weak), 1998, 2000, and 2008. In general, the ITCZ during La Niña is smaller than average in area, although that was not the case in 2000.

The IR-only labels were used to investigate climatic trends in the ITCZ from 1980 to 2009. The IR satellite record had some data gaps in the early 1980s and, although the statistical model is able to handle gaps in data, diagnosis is less reliable if data are not available for several sequential time steps. Therefore IR images where greater than 20% of the image was missing were removed unless the images on each side in time were more complete. This led to less information going into the data series in 1980–91

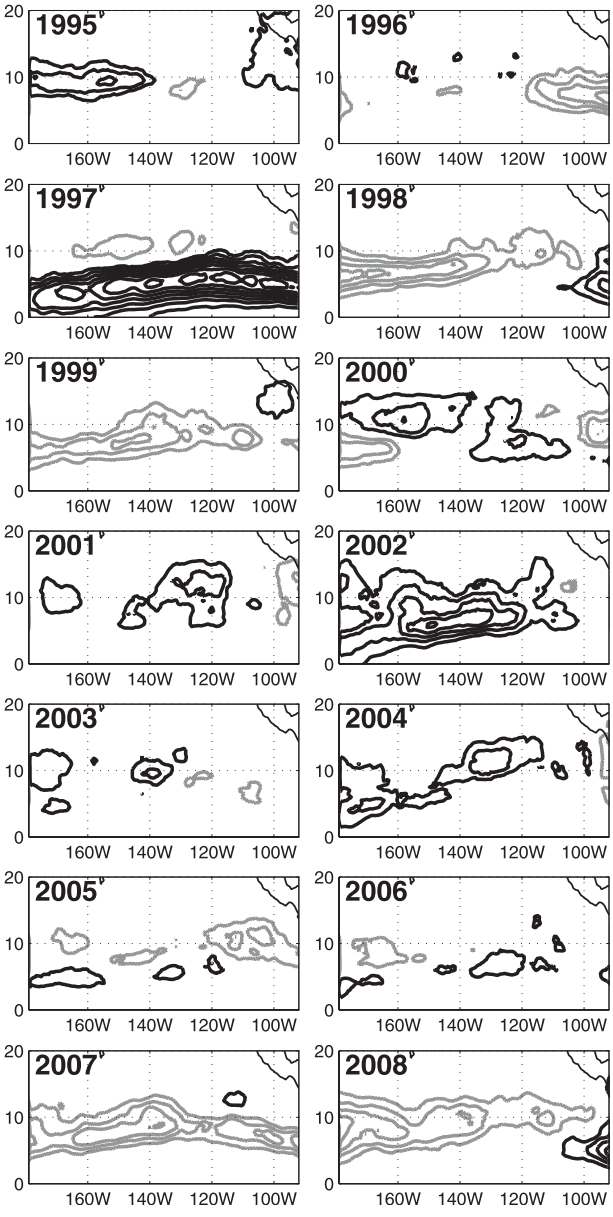

FIG. 9. Location anomalies from 1995 to 2008 using all-data labels. Black (gray) lines are positive (negative) anomalies. Each line represents a 10-day anomaly in mean occurrence over the season May–Oct.

when on average 2 days per month were missing or incomplete in the satellite record.

Figure 9 was reproduced using IR-only labels to examine location trends from 1980 to 2009. As mentioned at the beginning of the section, the area of the ITCZ was reduced in the 120°–150°W region owing to lower cloud not being identified in the IR. Again, there were positive anomalies in location in El Niño years and mostly negative in La Niña years. However, the figure is qualitatively similar to Fig. 9. Linear regression was performed to remove the influence of ENSO (in a similar manner to Vimont et al. 2001) on the mean annual ITCZ location so that climatic trends could be studied. The years when the multivariate ENSO index (Wolter and Timlin 1998) was greater than 0.5 were selected and the anomalies from mean location combined and weighted by ENSO index to produce a composite picture of typical El Niño year ITCZ location anomaly. This typical composite could then be removed from the El Niño years (weighted by the strength of ENSO). The same was performed for La Niña years when the ENSO index was less than −0.5. By this means, the influence of ENSO was removed or reduced, and any climatic shifts in location of the ITCZ not due to ENSO should be revealed. Several analyses were done on the resulting data and no significant or consistent trends in location were found. However, there is much variability from year to year, even after removal of the ENSO signature.

In addition to this analysis, histograms of latitudinal distribution each year were investigated and, again, no climatic shifts in the position of the ITCZ could be detected. This suggests that any trends in ITCZ location over the 30 years (that are not due to ENSO), are quite small or nonexistent.

Figure 10a shows the mean area of ITCZ from the IR-only labels for May to October each year, alongside the ENSO index. It should be noted that the time axis (x axis) is not continuous. Only the months of May through October are depicted each year. The plot shows high correlation between the ENSO index and the area covered by ITCZ. The correlation coefficient is 0.67, confirming the influence of ENSO in determining ITCZ area. To determine the long-term trends in the area of ITCZ independent of ENSO, the ITCZ data were normalized to vary between 0 and 1 and the normalized ENSO index was removed. The resulting time series is shown in Fig. 10b. The plot shows that there is no overall trend in ITCZ area over the 30 years, but there is much variability in size from one year to the next.

6. Concluding remarks

Our method has identified the ITCZ as a distinct weather feature in the east Pacific in instantaneous satellite data. The dataset has revealed the seasonal evolution of the ITCZ in the east Pacific and its interannual variability. The location and area of the ITCZ varies significantly on interannual time scales and is highly correlated with the ENSO index. During El Niño years there were significant shifts in ITCZ location in tandem with shifts in ocean warm-pool regions and the average area of the ITCZ was greater than in ENSO-neutral years. Inspection of the 30-yr ITCZ dataset using IR-only labels, showed

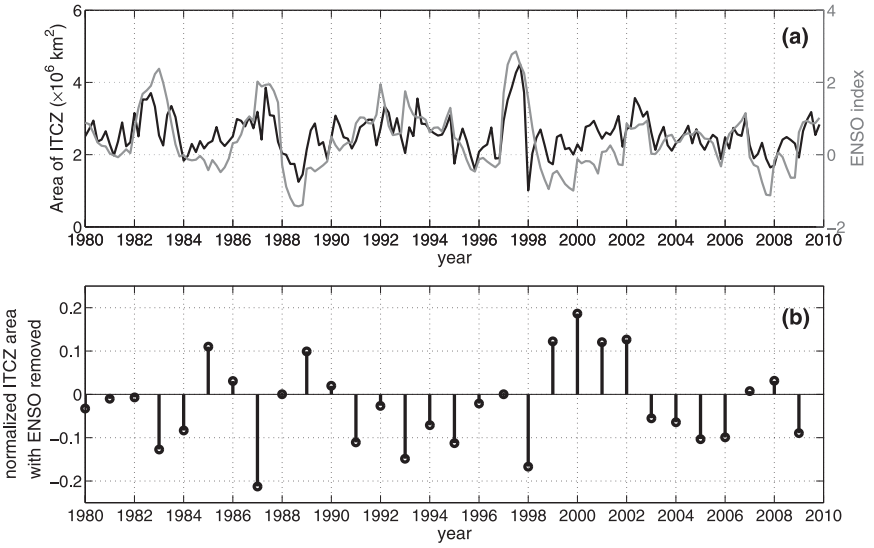

FIG. 10. (a) Mean area of ITCZ labels in black (monthly means May to Oct) shown alongside the ENSO index for the same months. (b) Area of mean annual ITCZ normalized with normalized ENSO index removed.

no consistent trend in the area covered by ITCZ or shifts in ITCZ location when the influence of ENSO was removed.

The ITCZ climatology of Waliser and Gautier (1993), which used cold cloud thresholding of the seasonal average to determine the signature of the mean convergence zone, described the ITCZ in the Pacific in a general sense. The mean latitude location was given as 8°N. This is in general agreement with our results, but moreover our results show the seasonal migration of the ITCZ and significant interannual variability. The ITCZ location is closely associated with SST. In the mean fields, the far eastern Pacific ITCZ is located to the south of the maximum SSTs, whereas in the central Pacific (150°W–180°), the ITCZ is located to the north of the maximum SSTs.

Our results show that there is a westward shift in ITCZ location throughout boreal summertime. The ITCZ is more concentrated in the east Pacific in May to July and more distributed to the central Pacific in September and October. The ITCZ area in the east Pacific peaks in late June and late August. The secondary maximum could be due to an increase in the number of tropical cyclones and other WPDs in August.

The application of statistical techniques in identifying the ITCZ has been shown to be successful. The labels automatically generated from the statistical model present a time-saving alternative to manual labeling. The model allows vast quantities of data to be analyzed for the presence of meteorological features at relatively little computational cost. The model uses a fast, portable code that can be run on any desktop: to identify the ITCZ for 6 months of 3-hourly data takes approximately 20 min on a parallelized computer using eight processors.

The use of the statistical model is ideal for this particular weather feature because of the ITCZ's persistent nature (rendering the use of the time component in the Markov random field highly appropriate), and the need to discard unrelated cloud features in the satellite images. Validation against manual labelers found that the statistical model was better equipped to identify coherent structures than thresholding techniques. Furthermore, thresholding of IR does not include low-cloud signatures in the ITCZ that are important for climate studies (Clement et al. 2009). The primary advantage of the MRF approach over other techniques is its accuracy in determining the ITCZ envelope. The model is able to take expert opinion, in the form of manual labels, and systematize the evaluation of the phenomenon over a long time period.

Here we have presented the seasonal and interannual characteristics of the ITCZ, but other uses of the data take advantage of the high temporal sampling of the new ITCZ labels (e.g., Bain et al. 2010). The 3-hourly time scale available in this dataset will allow investigations of internal ITCZ dynamical interactions as well as ITCZ relationships to different tropical features from the Madden–Julian oscillation to Kelvin waves. Future work could also experiment to improve the model, perhaps using different neighborhood structures, adding a time dimension to the location prior, or clarifying the impact of altering the strength of the β. With the success of the method there is also the possibility of applying the technique to the detection of the ITCZ in other regions of the world. The

model is thus far trained on manual labels of the ITCZ in the Pacific region but has the potential to be retrained for other global locations if manual labeling is carried out. Postprocessing can be tailored to the needs of any user. Another research possibility is to apply this MRF approach to tracking non-ITCZ features such as weather systems in the extratropical storm tracks. It also has the potential to be used in other fields such as oceanography to track ocean eddies or phytoplankton blooms.

Acknowledgments. The authors would like to acknowledge Ken Knapp at NOAA's National Climatic Data Center for providing all IR and VS data from the HURSAT-Basin archive. We would also like to thank Deborah Smith at Remote Sensing Systems for providing the TPW data from combined radar satellite instruments. We thank Mark DeMaria for valuable comments during this research and Yi-Hui Wang for help with labeling data. This research was supported by NSF Grant ATM0530926 and NOAA Grant NA09OAR4310132.

APPENDIX

Identification of Tropical Cyclones

Tropical cyclones produce large cloud signatures that can be embedded in or isolated from the main cloudy region of the ITCZ. It is therefore of interest to know what parts of the ITCZ labels may be associated with tropical cyclones.

The horizontal extent of tropical cyclones is most often distinguished by regions of sustained wind speed over 17 m s^{-1}; for example, as specified by the National Hurricane Center (more information available online at www.nhc.noaa.gov). However, higher wind speeds do not necessarily equate to the full extent of the cloud cover associated with a cyclone, and wind speeds are often only above tropical storm speed within a few degrees of the center of the eye. Therefore a second algorithm using the same IR satellite data as our statistical model was applied to identify the clouds belonging to tropical cyclones. Best-track data from the HURDAT dataset (more information available online at http://www.ncdc.noaa.gov/oa/rsad/hursat/) was used to establish the location of the eye of the storm. The data was linearly interpolated from the 6-h dataset to every 3 h. High clouds (<255 K) overlapping a nine-gridpoint neighborhood near the eye were selected, along with any intersecting low cloud (<270 K), but not nonintersecting high cloud. In this way if a cyclone were embedded in the ITCZ, the cloud belonging to the cyclone (only) was distinguished. A maximum radius was defined as 400 km, but this was rarely reached by the algorithm.

This method was seen as a balance between using more complex methods of pattern recognition and simpler methods such as defining a generic circle around all cyclone eyes as tropical cyclone cloud. The algorithm was able to adapt when cyclone-associated cloud was unusually positioned or the cyclone was embedded within a larger cloudy region. The inferred cyclone data is used in section 5a, and may be used for future investigations on interactions between cyclones and the ITCZ.

REFERENCES

Bain, C. L., G. Magnusdottir, P. Smyth, and H. Stern, 2010: Diurnal cycle of the Intertropical Convergence Zone in the east Pacific. *J. Geophys. Res.,* **115,** D23116, doi:10.1029/2010JD014835.

Clement, A., R. Burgerman, and J. Norris, 2009: Observational and model evidence for positive low-level cloud feedback. *Science,* **325,** 460–464.

Ferreira, R. N., and W. H. Schubert, 1997: Barotropic aspects of ITCZ breakdown. *J. Atmos. Sci.,* **54,** 261–285.

Garcia, O., 1985: Atlas of highly reflective clouds for the global tropics: 1971–1983. U.S. Dept. of Commerce, NOAA, Environmental Research Lab., Tech. Rep. NTIS PB-87129169, 365 pp.

Gelman, A., J. B. Carlin, H. S. Stern, and D. B. Rubin, 2003: *Bayesian Data Analysis*. Chapman and Hall/CRC, 696 pp.

Geman, S., and D. Geman, 1984: Stochastic relaxation, Gibbs distribution and the Bayesian restoration of images. *IEEE Trans. Pattern Anal. Mach. Intell.,* **6,** 721–741.

Gilks, W., S. Richardson, and D. Spiegelhalter, 1993: *Markov Chain Monte Carlo in Practice*. Chapman and Hall/CRC, 512 pp.

Gu, G., and C. Zhang, 2002a: Cloud components of the ITCZ. *J. Geophys. Res.,* **107,** 4565, doi:10.1029/2002JD002089.

——, and ——, 2002b: Westward-propagating synoptic-scale disturbances and the ITCZ. *J. Atmos. Sci.,* **59,** 1062–1075.

Kindermann, R., and J. L. Snell, 1980: *Markov Random Fields and Their Application*. American Mathematical Society, 143 pp.

Knapp, K. R., 2008: Calibration assessment of ISCCP geostationary infrared observations using HIRS. *J. Atmos. Oceanic Technol.,* **25,** 183–195.

——, and J. P. Kossin, 2007: New global tropical cyclone data from ISCCP B1 geostationary satellite observations. *J. Appl. Remote Sensing,* **1,** 013505, doi:10.1117/1.2712816.

Li, S. Z., 1995: *Markov Random Field Models in Computer Vision.* Lecture Notes in Computer Science. Springer, 264 pp.

Magnusdottir, G., and C.-C. Wang, 2008: Intertropical convergence zones during the active season in daily data. *J. Atmos. Sci.,* **65,** 2425–2436.

McPhaden, M. J., 1999: Genesis and evolution of the 1997-98 El Nino. *Science,* **283,** 950–954.

Mitchell, T. P., and J. M. Wallace, 1992: The annual cycle in equatorial convection and sea surface temperature. *J. Climate,* **5,** 1140–1156.

Raymond, D. J., G. B. Raga, C. S. Bretherton, J. Molinari, C. Lopez-Carrillo, and Z. Fuchs, 2003: Convective forcing in the intertropical convergence zone of the eastern Pacific. *J. Atmos. Sci.,* **60,** 2064–2082.

——, C. S. Bretherton, and J. Molinari, 2006: Dynamics of the intertropical convergence zone of the east Pacific. *J. Atmos. Sci.,* **63,** 582–597.

Scharenbroich, L., G. Magnusdottir, P. Smyth, H. Stern, and C.-C. Wang, 2010: A Bayesian framework for storm tracking

using a hidden state representation. *Mon. Wea. Rev.,* **138,** 2132–2148.

Serra, Y., and R. A. Houze, 2002: Observations of variability on synoptic timescales in the east Pacific ITCZ. *J. Atmos. Sci.,* **59,** 1723–1743.

Smyth, P., 1997: Belief networks, hidden Markov models, and Markov random fields: A unified view. *Pattern Recognit. Lett.,* **18,** 1261–1268.

Thorncroft, C., and K. Hodges, 2001: African easterly wave variability and its relationship to Atlantic tropical cyclone activity. *J. Climate,* **14,** 1166–1179.

Vimont, D. J., D. S. Battisti, and A. C. Hirst, 2001: Footprinting: A seasonal connection between the tropics and mid-latitudes. *Geophys. Res. Lett.,* **28,** 3923–3926.

Waliser, D. E., and C. Gautier, 1993: A satellite-derived climatology of the ITCZ. *J. Climate,* **6,** 2162–2174.

Wang, C.-C., and G. Magnusdottir, 2005: ITCZ breakdown in the three-dimensional flows. *J. Atmos. Sci.,* **62,** 1497–1512.

——, and ——, 2006: The ITCZ in the central and eastern Pacific on synoptic time scales. *Mon. Wea. Rev.,* **134,** 1405–1421.

——, C. Chou, and W.-L. Lee, 2010: Breakdown and reformation of the intertropical convergence zone in a moist atmosphere. *J. Atmos. Sci.,* **67,** 1247–1260.

Wolter, K., and M. S. Timlin, 1998: Measuring the strength of ENSO events—How does 1997/98 rank? *Weather,* **53,** 315–324.

Zhang, C., M. McGauley, and N. A. Bond, 2004: Shallow meridional circulation in the tropical eastern Pacific. *J. Climate,* **17,** 133–139.