

UC Irvine

UC Irvine Previously Published Works

Title

Automated analysis of the temporal behavior of the double Intertropical Convergence Zone over the east Pacific

Permalink

<https://escholarship.org/uc/item/6cr179x3>

Authors

Henke, Daniel
Smyth, Padhraic
Haffke, Colene
[et al.](#)

Publication Date

2012-08-01

DOI

10.1016/j.rse.2012.03.022

Copyright Information

This work is made available under the terms of a Creative Commons Attribution License, available at <https://creativecommons.org/licenses/by/4.0/>

Peer reviewed



Automated analysis of the temporal behavior of the double Intertropical Convergence Zone over the east Pacific

Daniel Henke ^{a,*}, Padhraic Smyth ^b, Colene Haffke ^c, Gudrun Magnusdottir ^c

^a Remote Sensing Laboratories, Department of Geography, University of Zurich, Switzerland

^b Department of Computer Science, University of California, Irvine, United States

^c Department of Earth System Science, University of California, Irvine, United States

ARTICLE INFO

Article history:

Received 5 September 2011

Received in revised form 13 March 2012

Accepted 17 March 2012

Available online 16 May 2012

Keywords:

Double ITCZ
Cloud tracking
Time series analysis
Hidden state space model
Hybrid HMM
SVM
Semi HMM
Semi-supervised
Image processing

ABSTRACT

This paper presents new methods for an automated analysis of the double InterTropical Convergence Zone (dITCZ) phenomena on a daily time scale over the east Pacific. Long-term Geostationary Operational Environmental Satellite (GOES) visible and infrared data are used to spatially identify and segment the convection zones over the east Pacific basin on both sides of the equator and to track the temporal variability of the ITCZ, specifically to identify cases of dITCZs, northern or southern ITCZ, or non-presence events. For the segmentation approach, image processing techniques are developed to extract information about the spatial features of the ITCZ in both hemispheres for each satellite image. These features serve as input to a temporal classification algorithm that is based on a combination of hidden semi Markov model (HsMM) and support vector machine (SVM) methods. The performance of the proposed method is competitive with human experts and the methodology can thus be used to conduct an in-depth analysis of the dITCZ. Such an analysis could provide precise information for refining existing weather and climate models over the sparsely observed east Pacific where the dITCZ is greatly over-represented in most models.

© 2012 Elsevier Inc. All rights reserved.

1. Introduction

The InterTropical Convergence Zone (ITCZ) is a zonally elongated, rather narrow zone of convection over the tropical oceans, playing a key role in the general circulation of the atmosphere. Understanding its variability is essential for improving global climate models. The ITCZ is frequently observed north of the equator in the east Pacific (hereafter referred to as the northern ITCZ (nITCZ)) especially during the boreal summer half-year. Even in winter, the ITCZ may be detected north of the equator as seen in cloud fields. Only in the boreal spring (especially March to April) is the ITCZ also observed south of the equator, which we will refer to as the southern ITCZ (sITCZ). At times this leads to two convergence zones, strong enough to be revealed in cloud fields, simultaneously existing on both sides of the equator in the tropical east Pacific. This phenomenon has been named the double ITCZ (dITCZ) (Hubert et al., 1969). Fig. 1 illustrates representative images from the visible channel on geostationary satellites for all four possible combinations of east Pacific ITCZ occurrences in the boreal spring, i.e. (a) dITCZ, (b) nITCZ, (c) sITCZ and (d) with no ITCZ signal present on either side of the equator. In this paper we reserve the symbol dITCZ to describe the occurrence of simultaneous convection

zones on both sides of the equator in the east Pacific in instantaneous data. Most previous studies have examined time-averaged and/or area-averaged fields and upon detecting convection both north and south of the equator refer to the phenomenon as a double ITCZ – this is different from our definition of dITCZ above.

Most current global climate models grossly over-represent the dITCZ during the boreal spring (active season). Some of the models even show a dITCZ outside of its active season. This leads to serious model misrepresentations of surface energy fluxes and radiative heating in the east Pacific with global consequences. Previous studies of the double ITCZ have focused on monthly (e.g. Liu & Xie, 2002; Zhang, 2001) or at most weekly time scales (e.g. Gu et al., 2005; Lietzke et al., 2001). Magnusdottir and Wang (2008) examined double ITCZ occurrence in daily data but used reanalysis data that by design include numerical model assumptions that may not be valid in the tropics. Masunaga and L'Ecuyer (2010) examined seasonal evolution of the southern branch of the ITCZ in eight years of daily satellite data (2000–2007) and the associated seasonal evolution of surface energy fluxes and the ocean mixed layer.

We are interested in characterizing the occurrence of each of the configurations (or states) of ITCZ on both sides of the equator in the east Pacific on the daily time scale. This requires an approach that relies only on instantaneous meteorological satellite images, avoiding potential corruption from meteorological analyses that may misrepresent the state of the tropical atmosphere in data sparse regions, such

* Corresponding author. Tel.: +41 44 6356524.

E-mail address: daniel.henke@geo.uzh.ch (D. Henke).

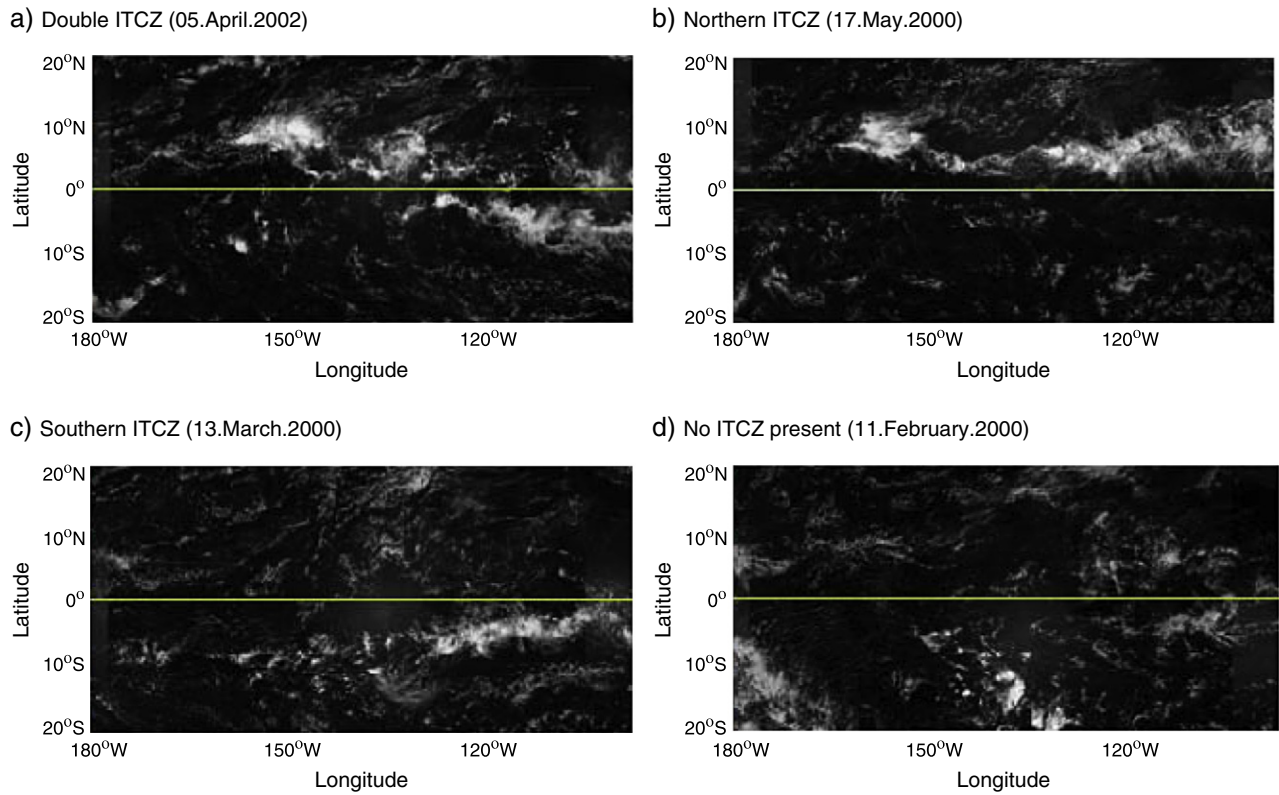


Fig. 1. Examples of the four possible ITCZ states in GOES VS images. The green lines indicate the equator.

as the tropical east Pacific. In addition, the approach should work in an automated manner to the extent possible in order to handle large data sets for long-term analyses that are impractical for manual annotation.

The algorithm we propose for this task consists of two major parts: first, we spatially localize the ITCZ in both hemispheres in single images (detection); and second, we use time-series analysis to identify different ITCZ states for later statistical analyses. For the first task, we rely on finding a backbone path to indicate the elongated ITCZ cloud band and use a subsequent region-growing algorithm to extract the connected ITCZ cloud field. An alternative approach to this unsupervised methodology would be to use spatio-temporal Markov random fields as described in Bain et al. (2011) – however, this requires time-consuming labeling of images at the pixel level, whereas in the approach proposed here no labeling is required for the spatial segmentation.

For the second task, features extracted from the spatial segmentation are used to classify each image into one of four discrete states (dITCZ, nITCZ, sITCZ and no presence). We make use of an extended hidden semi Markov model/support vector machine (HsMM/SVM) approach, leveraging HMM modeling techniques to integrate information over time (Rabiner, 1989), adapting hybrid HMM methods that traditionally have been applied to speech recognition (Boullard & Morgan, 1994).

Our results in this paper demonstrate that the proposed method can extract extensive information about the ITCZ phenomenon – about its current state as well as characteristics such as extent and location – at a high temporal resolution and in a fully automatic manner. The method is competitive in accuracy when compared to human experts and it significantly reduces the amount of human effort involved in annotation since only an initial set of training images need to be labeled and not the whole data set.

In the following, we first briefly describe the satellite data that we use and the ground truth acquisition by human experts. Then the method is described in detail. Beginning with localization of the

ITCZ, we will focus on the temporal classification of the dITCZ states. We evaluate the performance of the proposed algorithm and present first basic climate statistics based on a five-season sample case. Finally, we conclude with potential future research directions.

2. Methods

2.1. Dataset and ground truth

The visible (VS) and infrared (IR) data of the GridSat data set from Knapp et al. (2011) are used in this paper.¹ In GridSat, raw data from multiple geostationary satellites are remapped to a standard map projection, recalibrated to optimize temporal homogeneity (especially optimized for the IR data), and stitched together from different satellite instruments to provide a standardized data set for the years 1980 to present. The spatial resolution of the data is approximately 8 km which allows accurate identification of large-scale ITCZ phenomena. VS images are available daily at 21:00 GMT (daylight) while IR images are available every 3 hours since they are independent of the position of the sun. High reflectivity values in VS images correspond to optically thick cloud cover and low IR temperature corresponds to cold cloud tops or cloud tops high in the troposphere. These characteristics are useful indicators for the presence of the ITCZ which, in general, is associated with deep as well as shallow convection (see also Bain et al., 2011). For detecting the dITCZ, nITCZ, and sITCZ states, we restrict the area of interest to longitudinal and latitudinal ranges of 180° W to 90° W and 20° N to 20° S and the period of potential occurrence to February until May of each year. In subsequent analysis of the

¹ The files stored in GRISAT-B1 beta netCDF-4 format were provided by Dr. Ken Knapp and are precisely described in Knapp et al. (2011). They were generated for the boreal spring (i.e. February–May) of the years 2000–2004 and the specific region of interest (i.e. east Pacific).

images we rescaled the [0...1] reflectivity values of the VS data and the temperature-normalized IR data to 256-integer images.

Human experts (atmospheric scientists familiar with the ITCZ process) manually generated image labels using a graphical user interface. To assist the human experts three different satellite fields (VS, IR and Total Precipitable Water (TPW)) were displayed in the interface. TPW was obtained from Remote Sensing Systems and is a composite from various microwave sensors. Two atmospheric scientists (Experts I and II) each independently labeled two seasons of data (2000 and 2002), and two more experts (Expert III and IV) each labeled a sequence for March 2000. The sequence labeled by Expert I was used for training and testing of the model; the accuracy of Expert I was measured using the labels from Experts II, III and IV. The scientists assigned each day to one of four stage tags, where the mutually exclusive stage tags were (a) dITCZ, (b) nITCZ, (c) sITCZ, or (d) a non-presence event. The purpose of collecting labels in this manner was to provide ground truth data for training pattern recognition algorithms, as well as for systematically evaluating the quality of such algorithms by comparing algorithm and human accuracy on held-out data.

2.2. Locating the ITCZ – the backbone algorithm

The first challenge for automatic tracking of the dITCZ phenomena is to locate the ITCZ position (if present) and its spatial extent, on both sides of the equator. As a preprocessing step, the data are filtered by a standard Gaussian spatial filter. A latitude-dependent weighting (a mixture of two smooth Gaussian functions with maxima at 8° N and 8° S, see Appendix A) is used to add prior knowledge about the most likely ITCZ location in order to bias the path-finding algorithm toward these regions. Thereafter, we estimate the location of the ITCZ by finding a path (or “backbone”) that is optimal in the sense that it maximizes the intensity value of the summed pixels from the west to the east. Dynamic programming is used to compute the optimal path as outlined below.

Let I be a two-dimensional $N \times M$ array of intensity values corresponding to the satellite image with N referring to the number of columns and M to the number of rows. Let P be the matrix storing the estimated backbone path from left to right and P^A an auxiliary variable containing the path positions:

$$\begin{aligned} \text{Initiation: } & P(1, m) = I(1, m) \quad \forall m : 1 \leq m \leq M \\ \text{Recursion: } & P(n, m) = I(n, m) + \max_{m-1 < k < m+1} P(n-1, k) \quad \forall n : 2 \leq n \leq N; \forall m : 1 \leq m \leq M \\ & P^A(n, m) = \arg \max_{m-1 < k < m+1} P(n-1, k) \quad \forall n : 2 \leq n \leq N; \forall m : 1 \leq m \leq M \end{aligned} \quad (1.1)$$

To estimate the backbone path positions P^O the maximum of P in column N is chosen and backtracked using the auxiliary variable P^A .

$$\begin{aligned} \text{Initiation: } & P^O(N) = \arg \max_{1 < k < M} P(N, k) \\ \text{Recursion: } & P^O(n) = P^A(n+1, P^O(n+1)) \quad \text{with } n = N-1, N-2, \dots, 1 \end{aligned} \quad (1.2)$$

We apply this algorithm to both sides of the equator, i.e. 20° N to 0° for the northern ITCZ and 0° to 20° S for the southern ITCZ. Fig. 2 shows a typical result from this algorithm, illustrating the estimated backbone paths (and corresponding ITCZ locations) obtained from VS data.

Given the estimated path we extract the spatial extent of the ITCZ as follows. Morphological reconstruction by dilation is used to find a spatial segmentation. To avoid the computationally-intensive process of iterative geodesic dilation we use a downhill filter which relies on a random access queue and is initialized via seed points defined by the backbone path – this approach was suggested by Robinson and

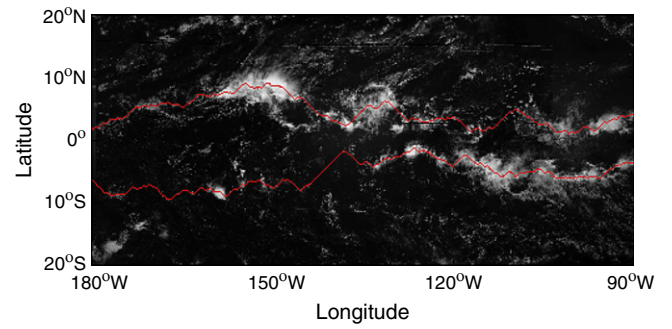


Fig. 2. The red line indicates the location of the estimated backbone path (in both the northern and southern hemispheres), and correspondingly, the estimate of the dITCZ location. From VS data.

Whelan (2004) in the context of medical image analysis. The spatial segmentation algorithm is as follows:

Initialize:

The mask image C is defined by the original satellite image

$$I(p) = \begin{cases} C(p) & \text{if } p \text{ is a pixel of the backbone path} \\ 0 & \text{otherwise} \end{cases} \quad (2.1)$$

The access list L is defined as: $\forall p : I(p) \neq 0$: Append p to $L(I(p))$

The auxiliary variable m is set to the maximum value of I .

Loop over all $n = m \dots 1$:

While $L(n) \neq \{\}$ do

Take first element p out of $L(n)$ and do $\forall q \in \text{Neighborhood}(p)$ and q not already visited:

$I(q) = \min(n, C(q))$ and append q to $L(I(q))$

After the loop is completed the pixels in I belonging to the ITCZ signal are preserved while non-ITCZ pixels are suppressed (see Appendix A).

This type of region-growing algorithm tends to preserve the main cloud band of the ITCZ while simultaneously suppressing cloud formations which are not connected to the ITCZ complex. The resulting image I is thresholded to produce a mask – this mask corresponds to our estimate of the spatial location and extent of the ITCZ (see Appendix A). The method is not particularly sensitive to the choice of threshold as long as it is above a minimum value – a value of 60 (using a scaled VS image with a reflectivity range [0...255]) was used in the experiments. Finally, we use morphologic open and close operations to connect nearby segments and to filter out small regions. For the close operation we use an elliptically shaped mask with an east–west orientation of the principal axis to favor elongated east–west oriented cloud bands (see Appendix A for details). The method is applicable to both the VS and inverted IR images.² Fig. 3 illustrates the estimated mask for the example from Fig. 2.

2.3. Time series analysis of the dITCZ phenomena

The second primary challenge in analyzing the dITCZ phenomena is to track and classify ITCZ presence over time in both the northern and southern hemisphere. Our approach to this temporal classification problem is to first estimate features from the satellite images, and to then use these features as observations for a hidden state-space model, from which we can estimate the most probable state sequence, i.e. presence or non-presence over time of ITCZ north and south of

² When we display or refer to IR images in what follows we use inverted IR images (i.e., bright pixels correspond to lower intensity values) since this improves readability and understanding.

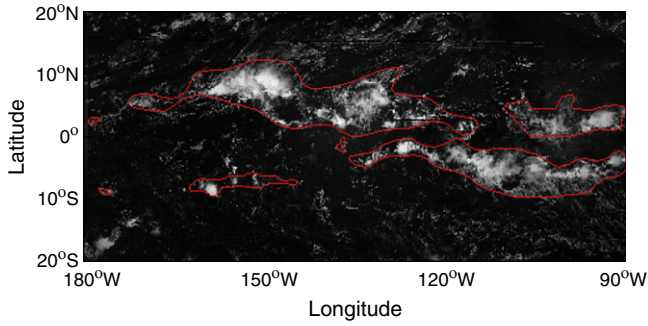


Fig. 3. Result from automatic spatial segmentation of the ITCZ corresponding to the image in Fig. 2.

the equator. As mentioned earlier, and as illustrated in Fig. 1, we distinguish between four categorical and mutually-exclusive states:

- The dITCZ state where an ITCZ signal is visible (consisting of at least a roughly 90° longitudinal stretch of a connected cloud band³) on both sides of the equator;
- The northern state, where only one ITCZ is formed in the northern hemisphere (nITCZ);
- The southern state, where only one ITCZ is formed in the southern hemisphere (sITCZ);
- The state of non-presence, i.e. no significant ITCZ signal in the area of interest.

In preliminary experiments (not described here) we found that the 4-state model described in this paper outperformed two independent 2-state models that analyze the northern and southern ITCZ separately. This is suggestive of correlations in the dITCZ phenomenon which we will comment on again later in the paper. In what follows below, we will first describe the feature extraction and then give a detailed overview of the temporal classification methods.

2.3.1. Feature extraction

We use a large number of different image-based features in our temporal classification approach (34 for VS and 37 for IR) – here we present only an overview – the complete list of features is defined in Appendix B. The features can be sub-divided into three different classes: those based on the backbone path, those based on the segmentation mask, and those based on general image statistics. Features are calculated separately for the northern part (20° N to 0°), the southern part (0° to 20° S), or in some cases for the complete image (20° N to 20° S). The features are calculated in the same manner for the VS and IR images, the only difference being in the threshold used for the segmentation mask.

2.3.1.1. Features based on the location of the backbone path. We extract features from the backbone path such as the slope or spatial variance of the path with respect to a linear best-fit regression. We find that this variance is typically lower when the ITCZ is present due to the typical elongated linear form of an ITCZ. Additionally, we derive statistics from the pixel intensities along the backbone path. For example, the mean of the intensity values as well as the distance of local minima from each other tend to be higher when the ITCZ is present since the ITCZ tends to be associated with elongated high-intensity cloud cover.

³ Note: A 90° longitudinal connected cloud band is our working definition of the ITCZ being present but the interpretation of a “connected cloud band” is subjective. This subjectivity naturally leads to variations among the labels assigned by the atmospheric scientists.

2.3.1.2. Shape-based features. Shape-based features are derived from the binary mask that is calculated by morphological reconstruction described in Section 2.2. We use border tracing as described in Klette and Rosenfeld (2004) to extract all connected objects. Subsequent features that can be computed include for example the width of the object that has the largest east–west extent. This width is often quite large when the ITCZ is present relative to images without the ITCZ. For the object with the widest east–west extent additional characteristics such as area, perimeter, center of mass and compactness are also computed and used as features.

Shape-based features are also derived from Fourier descriptors (Arbter et al., 1990) computed on the border. The Fourier coefficient c_0 defines the balance point of the border line. Using the Fourier coefficient c_1 , and the corresponding negative coefficient c_{-1} , an ellipse is fitted to the extracted object, as described in Appendix B. The major axis (which for the ITCZ is related to the maximal east–west width), the minor axis, and its ratio and orientation all provide useful features. The higher order Fourier coefficients are indicative of the smoothness of the object’s border and thus the energy spectrum of the Fourier coefficients also serves as an additional feature.

2.3.1.3. Features based on general image statistics. As basic features we calculate the mean intensity of the complete image, for both the northern and southern parts of the image separately, as well as for the areas 12.5° N to 2.5° N and 2.5° S to 12.5° S (as defined for the double ITCZ phenomenon over the western Pacific in Chen et al., 2008). Central and invariant moments of the intensity distribution can also provide useful information and a number of such moments are used as features.

2.3.2. Temporal classification of the dITCZ phenomena with hidden state space models

Our aim is to classify each time snapshot into one of four states, conditioned on the observed features. Since the features are somewhat noisy, we use a temporal classifier to model the persistence of the dITCZ states. Specifically, we use a hidden Markov model (HMM) to incorporate knowledge about the temporal behavior of the dITCZ phenomena. In an HMM the probability distribution over the hidden state of the system at a particular time depends only on (a) the previous state of the system (the Markov property) and (b) the current observation.

As described in Rabiner (1989) an HMM is formally defined as

- (1) a set of N discrete states $\{q_1, q_2, \dots, q_N\}$
- (2) a state transition matrix $\mathbf{A} = \begin{pmatrix} a_{11} & \dots & a_{1N} \\ \vdots & & \vdots \\ a_{N1} & \dots & a_{NN} \end{pmatrix}$ with $a_{ij} = P(q_t = j | q_{t-1} = i)$ being the probability that state j follows state i under the conditions $a_{ij} \geq 0$ and $\forall i: \sum_{j=1}^N a_{ij} = 1$.
- (3) a set of emission probability distributions or densities: $B = \{b_1, \dots, b_n\}$; where $b_i(\mathbf{x}) = P(\mathbf{o}_t = \mathbf{x} | q_t = i)$ is the probability of generating observation \mathbf{x} in state i .
- (4) a feature or observation space O which is the domain of the b_i ’s.
- (5) an initial probability distribution $\boldsymbol{\pi}$ where $\pi(i)$ is the probability of starting in state i .

For modeling the dITCZ phenomena the states consist of {dITCZ, nITCZ, sITCZ, non-presence} and the observations \mathbf{O} correspond to the set of features described earlier.

In a typical application of the HMM approach, the model parameters $\boldsymbol{\theta} = (\mathbf{A}, B, \boldsymbol{\pi})$ are learned in an unsupervised manner by maximizing the probability $P(\mathbf{O} | \boldsymbol{\theta})$ for a given realization $\mathbf{O} = \{\mathbf{o}_1, \dots, \mathbf{o}_T\}$ (the training data) via the Baum–Welch algorithm, which is a specific case of the more general expectation maximization (EM) procedure (e.g., see Rabiner, 1989). For a new set of images whose states

are unknown, given the features of the image sequence \mathbf{O} and the trained model parameters θ , the computation of the optimal state sequence $\mathbf{Q} = \{q_1, q_2, \dots, q_T\}$ – here, the temporal tracking of the dITCZ phenomena – is carried out using the Viterbi algorithm (based on dynamic programming similar to the optimal path algorithm in Eqs. (1.1)–(1.2) of Section 2.2) by maximizing the probability $P(\mathbf{Q}|\mathbf{O}, \theta)$.

To improve the performance of the model for the purpose of dITCZ tracking we make three changes to the standard HMM framework. These modifications have been investigated independently in the past and proven to be useful in general but to our knowledge have never been used in combination:

We begin by relaxing the strict Markov property for the transition matrix \mathbf{A} by introducing a time-dependent self-transition function to the transition matrix, also known as a hidden semi-Markov model (HsMM) (e.g. Murphy, 2002). This modification allows for explicit modeling of the distribution of state durations.

Second, instead of modeling the emission probabilities B via a conditional (generative) model of the probability of the observations given the states (e.g., via a Gaussian distribution) we instead model the conditional distribution of the states given the observations using support vector machines (SVMs). The latter approach is generally less sensitive to parametric distributional assumptions than the former, particularly in high-dimensional observation spaces. The general approach is discussed for example in Bouldard and Morgan (1994) – here we specifically use SVMs and logistic regression (as in Platt, 1999) to model the conditional distribution of states given observations, as described later in this paper.

Finally, since training on known expert labels can improve the model's performance significantly (compared to unsupervised learning without any labeled data), we replace unsupervised parameter estimation with both supervised (all labels known) or semi-supervised (partly known and partly unknown labels) learning, as described for example by Zhong (2005).

2.3.2.1. Relaxing the strict Markov property: the HsMM. The standard HMM implies a geometric distribution for the state transition probability due to the constant self-transition coefficients (the a_{ii} 's in the transition matrix \mathbf{A}). The state duration probability density $p_i(d)$ for staying in state i for exactly d time-steps is

$$p_i(d) = (a_{ii})^{d-1} * (1 - a_{ii}) \quad (3)$$

This geometric distribution is inconsistent with prior knowledge about ITCZ persistence. For example, in Wang and Magnusdottir (2006), where the durations of northern ITCZs in the summer months were manually identified for the years 1999–2003, typical durations were found to be 9 days in length and non-geometric in terms of distribution. Furthermore, based on the human-generated labels used in this present study, we found that the state duration for the ITCZ in either hemisphere tends to have a mode that is centered between 5 and 15 days, rather than at 0 days as required by a geometric distribution. For these reasons we model the state duration explicitly by introducing a time-dependent component in the state transition matrix \mathbf{A} .

Let L be a contiguous (in time) labeled training sequence obtained from a human labeler. We define $d_k(i)$ as the duration of the k th connected occurrence of state i in L . The state duration distribution $D_i(t)$ over time for each state can be calculated as

$$D_i(t) = \sum_k 1(d_k(i) \leq t);$$

$$\text{with } 1(d_k(i) \leq t) = \begin{cases} 1 & \text{if } d_k(i) \leq t \\ 0 & \text{else} \end{cases} \quad \forall t: 1 \leq t \leq \max(d_k(i)) \quad (4.1)$$

We can approximate the state duration probability by fitting a logistic regression model of the following form:

$$D_i(t) = \frac{e^{\beta_0 + \beta_1 t}}{1 + e^{\beta_0 + \beta_1 t}} \quad (4.2)$$

where the β_k 's are estimated by maximum likelihood estimation (MLE). Since the training data for learning the state duration is relatively small, outliers can have a severe impact on the fitting of the β_k coefficients. To overcome this we use a scaled, interpolated and smoothed version of $D_i(t)$ as a look-up table for finite amounts of training data, recovering the logistic regression form for $t \gg \max(d_k(i))$. Additional details, including an example of such fitted duration models, are provided in Appendix C.

The constant a_{ij} 's in the transition matrix \mathbf{A} can now be replaced by the time-dependent $a_{ij}(t)$'s which are defined as the scaled, interpolated and smoothed version of the $D_i(t)$ values described above. To satisfy the criteria $\forall i: \sum_{j=1}^n a_{ij} = 1$ for each row of $A(t)$, we normalize the a_{ij} with $i \neq j$.

As mentioned earlier, the training data for learning the state duration is often relatively small since there are a limited number of ITCZ transition events per season. Thus, we found it useful to use a transition matrix that is a combination of purely dynamic and static transition matrices, i.e., $\mathbf{A}^{\text{new}}(t) = \lambda * \mathbf{A}^{\text{dynamic}}(t) + (1 - \lambda) * \mathbf{A}^{\text{static}}$ where λ is a weighting factor. $\mathbf{A}^{\text{static}}$ is more robust since it depends on fewer parameters than $\mathbf{A}^{\text{dynamic}}(t)$. As the number of training samples increases, $\mathbf{A}^{\text{dynamic}}(t)$ contributes more information, and the weighting factor λ can be increased. In our dITCZ experiments, for the supervised VS model with 1 image per day, and relatively little training data, we rely mainly on the static matrix with $\lambda = 0.1$. For the semi-supervised IR model with 8 images per day (thus 8 times as much training data) we use a purely dynamic transition matrix ($\lambda = 1$). Furthermore, as discussed below, $\lambda = 1$ is also necessary for robust semi-supervised learning. Finally, we use modified forms of the alpha-beta and Viterbi-algorithms by introducing an auxiliary variable that keeps track of the time spent in the current state making it possible to update the self-transition probability of $\mathbf{A}^{\text{dynamic}}(t)$ at a given time step t in the classification step.

2.3.2.2. Modeling the state-observation probabilities by SVMs: the HsMM/SVM hybrid. In the typical application of HMMs the b_i 's are defined as $b_i(\mathbf{x}) = P(\mathbf{o}_t = \mathbf{x} | q_t = i)$, the probability of generating observation \mathbf{x} in state i . However, common parametric forms for the b_i 's (such as the multivariate Gaussian distribution) are often not well-matched to real data distributions. As discussed earlier, the alternative parametrization we pursue here is to model $P(q_t = i | \mathbf{o}_t = \mathbf{x})$ by an SVM or by logistic regression. The necessary conditional probabilities for learning and classification for this parametrization can be derived using Bayes' rule:

$$P(\mathbf{x} | q_t = i) = \frac{P(q_t = i | \mathbf{x}) * P(\mathbf{x})}{P(q_t = i)} \quad (5)$$

$P(q_t = i)$ can be estimated via the labeled training set and $P(\mathbf{x})$ is a scaling factor since it is independent of the state s . It can be shown that the parameters of the H(s)MM are invariant with respect to scaling and thus one can convert any classifier which satisfies the maximum a posteriori (MAP) criterion (i.e., it estimates $P(q_t = i | \mathbf{x})$) into an estimator of the emission densities B of the HsMM (e.g. see Bouldard & Morgan, 1994).

When using SVMs some additional adjustments need to be made to implement this parametrization. Firstly, standard binary SVMs can only discriminate between two classes. Since we deal with four different classes (dITCZ, northern, southern, not present), we use "one-versus-all" SVMs for all classes, i.e., we use four binary classifiers,

one per state. Given our labeled state sequence L , we set up the l training data pairs $(\mathbf{x}_t, y_{i,t})$ for the SVM as follows:

$$y_{i,t} = \begin{cases} 1 & \text{if the state at time } t \text{ is } i \\ -1 & \text{otherwise} \end{cases} \quad (6.1)$$

with a corresponding feature vector \mathbf{x}_t .

With these l training data pairs the support vectors are learned by maximizing:

$$L(\alpha) = \sum_{k=1}^l \alpha_k - \frac{1}{2} \sum_{k=1}^l \sum_{m=1}^l y_k y_m \alpha_k \alpha_m \langle \mathbf{x}_k, \mathbf{x}_m \rangle \quad (6.2)$$

under the constraints $0 \leq \alpha_k \leq C$ and $\sum_{k=1}^l y_k \alpha_k = 0$

The cost parameter, C , controls the trade-off between allowing training errors and forcing rigid margins, and was set to a fixed default value $C = 1$ in all of our experiments.

Given the α 's, the hyperplane parameters of the SVM are defined as:

$$\mathbf{w} = \sum_{k=1}^l \alpha_k y_k \mathbf{x}_k \quad \text{and} \quad (6.3)$$

$$u = -\frac{1}{2} \left(\max_{k,y_k=-1} (\langle \mathbf{w}, \mathbf{x}_k \rangle) + \min_{k,y_k=+1} (\langle \mathbf{w}, \mathbf{x}_k \rangle) \right)$$

A second issue is that SVMs only calculate the distance of a new sample \mathbf{x}_{new} to the optimal hyperplane $\langle \mathbf{w}_i, \mathbf{x}_{new} \rangle + u_i$ but do not directly compute the probability $P(q=i|\mathbf{x})$. However, the distance to the optimal hyperplane implies a measurement of certainty about the classification decision (small distances imply low certainty, large distances imply high certainty). One can approximate the probability of $P(y_{i,t} = 1|\mathbf{x}_t)$ by a sigmoid function (Platt, 1999):

$$P(y_{i,t} = 1|\mathbf{x}_t) \approx \frac{1}{1 + e^{-\varepsilon_i (\langle \mathbf{w}_i, \mathbf{x}_t \rangle + u_i) + \gamma_i}} \quad (6.4)$$

where \mathbf{w}_i and u_i are the parameters learned by the SVM and ε_i and γ_i are estimated iteratively with respect to the training data. To fit the SVM models in our experiments we used a non-linear least square SVM (LS-SVM) with a radial basis function (RBF) kernel, as implemented in the LS-SVM MATLAB toolbox by De Brabanter et al. (2010). The width parameter of the RBF was set to 110 based on preliminary experiments with the dITCZ training data.

2.3.2.3. Supervised and semi-supervised learning approaches. For learning the model parameters θ we use both supervised and semi-supervised approaches depending on the particular data set used for the classification. Both data sets, VS and IR, have their advantages and disadvantages. The VS imagery (which is also the main source for the atmospheric scientists to visually detect the ITCZ signal) achieves the better results for classifying the dITCZ phenomena. One of the main reasons for this is that the ITCZ phenomena in the east Pacific in spring (Feb–May) can be weak, and thus, there are numerous shallow ITCZ events which are only detectable in VS data but are not seen in IR images. Including IR features in the VS model downgrades the model's accuracy. These facts in general favor VS data over IR data. On the other hand, IR data are available at a higher temporal resolution (8 per day vs. only 1 per day for VS), and more importantly, the time series for analyzing the long-term behavior is more reliable for the IR data since VS data are less temporally homogeneous. Furthermore, the VS data prior to 1995 can be corrupt or non-existent as discussed in previous investigations (e.g. Bain et al., 2011).

These differences between IR and VS motivated our investigation of two different learning approaches. For the VS model with only daily images, we have labels for each image and thus, the model parameters θ can be learned in a completely supervised manner, i.e. we have an expert labeled sequence L (which yields by definition a

most probable (with probability 1) state sequence Q) and an associated observation sequence O . A and π can be estimated directly from L , and B can be trained by the SVMs using the (O, L) -tuples.

In contrast, the IR images are available every 3 hours, i.e. 8 per day. In principle, these additional images (7 of which are unlabeled) provide the opportunity to improve the model's performance via a semi-supervised learning approach as follows. We make the reasonable assumption that the state duration for a specific ITCZ phase spans at least 1 day. We enforce this by making the self-transition duration of Eq. (4.1) be 1 for the first 8 time steps (8 IR images per day): $\forall i, 0 < t < 8: D_i(t) = 1$ and we use a purely dynamic transition matrix, i.e. $\lambda = 1$. Given an observation sequence $O = \{o_1, o_2, \dots, o_N\}$ and the corresponding labels $L^{known} = \{l_{8*k+m}\}$ with $0 \leq 8*k \leq N - m, 0 < m < 8, m$ fixed, we wish to maximize $P(O|\theta)$ under the constraint that the state sequence $Q = \{q_1, q_2, \dots, q_N\}$ at times $8*k + m$ equals exactly the supervised labels L^{known} , i.e. $\forall k: l_{8*k+m} = q_{8*k+m}$. We first initialize our model parameters $\theta = (A, \pi, w, u)$ as follows:

- Train parameters w_i and u_i of the SVMs by using the labels L^{known} and the corresponding observation sub-set $O = \{o_{8*k+m}\}$ with $0 \leq 8*k \leq N - m, 0 < m < 8, m$ fixed as for the VS model (Eqs. 6.1–6.4).
- Define new labels $L^r = \theta$ for the missing labels via the following simple heuristic: between two labels l_{8*k+m} and $l_{8*(k+1)+m}$, where $l_{8*k+m} = l_{8*(k+1)+m}$, set all labels between, i.e. $l_{8*k+m+1}, l_{8*k+m+2}, \dots, l_{8*k+m+7}$, equal to l_{8*k+m} . If a state transition exists between two labels, i.e. $l_{8*k+m} \neq l_{8*(k+1)+m}$, we set all values $l_{8*k+m+1}, \dots, l_{8*k+m+4}$ equal to l_{8*k+m} and all labels $l_{8*k+m+5}, \dots, l_{8*k+m+7}$ equal to $l_{8*(k+1)+m}$:

$$\begin{aligned} \text{For } j = \{1, 2, 3, 4\} : l_{8*k+m+j} &= l_{8*k+m} \\ \text{For } j = \{5, 6, 7\} : l_{8*k+m+j} &= l_{8*(k+1)+m} \end{aligned} \quad (7)$$

- Finally, the dynamic transition matrix A , derived from L^O and Eqs. (4.1)–(4.2), as well as the prior distribution π , are learned in a manner equivalent to the VS model case, except for forcing the minimum self-transition time to be at least 8 steps.

We use a quasi-EM algorithm for semi-supervised learning of HsMM/SVM models on the IR data. The aim is to obtain the optimal model parameters θ^* by maximizing the marginal likelihood function $\mathcal{L}(\theta; O) = P(O|\theta) = \sum_Q P(O, Q|\theta)$ given the observation sequence O : $\theta^* = \arg \max_{\theta} P(O|\theta)$. For our semi-supervised approach we include the additional constraint that the state sequence Q has to traverse the given labels L^{known} . Since there is no closed analytic solution to the problem, we solve it iteratively. In the estimation (E) step of the EM algorithm, the expected value of the log likelihood function under the current estimate of the parameters θ^r is calculated with respect to the conditional distribution Q given O using the following auxiliary function: $Aux(\theta|\theta^r) = E_Q[\log P(O, Q|\theta)|O, \theta^r]$. For our model this is simply equivalent to computing the most probable state sequence Q^r given O and θ^r under the constraint that Q^r traverses L^{known} . In the maximization (M) step we seek the parameters maximizing the auxiliary function: $\theta^{r+1} = \arg \max_{\theta} Aux(\theta|\theta^r)$. Thus, the estimate of the parameters θ^{r+1} is updated according to the state sequence Q^r of the E-step and the observation sequence O .

A more detailed description of the algorithm can be found in Appendix D.

We repeat the EM algorithm until no changes between the labels at r and $r+1$ occur, i.e. $L^r = L^{r+1}$, and thus, convergence is reached. It is possible (although it did not happen in any of our experiments) that the algorithm might not converge – to avoid this one can define a maximum number of iterations and then select the maximum over all parameter estimates $\theta = \{\theta^0, \theta^1, \dots, \theta^R\}$, thus, guaranteeing that the semi-supervised case has at least as high a likelihood as the supervised IR model: $P(O|\theta^{semi-supervised}) \geq P(O|\theta^{supervised})$.

Since we use a semi-supervised SVM/HsMM hybrid approach the strict EM-criterion, that $P(O|\theta^{r+1}) \geq P(O|\theta^r)$, is not necessarily

Table 1
ITCZ State distribution of expert I.

	dITCZ	nITCZ	sITCZ	not present
Expert I (241 days labeled)	27.8% (67 days)	55.6% (134 days)	11.6% (28 days)	5.0% (12 days)

Table 2
Classification accuracy of human experts.

	Expert II	Expert III	Expert IV
Expert I	77.2% (241 common days)	80.7% (31 common days)	74.2% (31 common days)

guaranteed – but the method appears to work well empirically as illustrated below.

3. Experimental results

We applied the methods described in Section 2 to the VS and IR satellite data and discuss below the results in terms of model performance relative to human labeling. We also discuss a number of insights gained about the temporal behavior of the dITCZ phenomena.

3.1. Performance of the models

As described in Section 2.1, to train and evaluate the models, four atmospheric scientists (Experts I–IV) independently labeled a sequence of satellite images using VS, IR and TPW images of the area of interest for each day. Two of the scientists labeled two seasons (Feb–May of 2000 and 2002, 241 days total). The other two scientists labeled 31 days in March 2000. The labels of Expert I were used for training and validating the model. The three additional expert labels provided information about typical subjective variability across human experts for the ITCZ-labeling task.

Table 1 shows the distribution of the fraction of days per label for the four disjunctive dITCZ states, using the labels of Expert I. We see that the nITCZ state dominates (55.6% of days).

Table 2 shows the classification accuracy of the labels of Experts II, III, and IV, relative to Expert I, i.e., treating Expert I's labels as ground truth. The accuracy ranges from 74% to 81%, indicating that there is a relatively high degree of subjectivity for this task. Measuring the inter-labeler accuracy is useful because it allows us to calibrate an automated algorithm's performance relative to human performance on the same task.

To evaluate the performance of our models, we divided the images (and labels from Expert 1) into a training and test set of equal size, such that the distribution of state occurrences in the training set and test set is approximately the same. The training set used for the statistics in the paper comprises 121 days (dITCZ: 37 (30.6%), nITCZ: 60 (49.6%), sITCZ: 17 (14.0%), no ITCZ: 7 (5.8%)) the test set 120 days (dITCZ: 30 (25.0%), nITCZ: 74 (61.7%), sITCZ: 11 (9.2%), no ITCZ: 5 (4.2%).⁴

The performance of the VS model on the test data is shown in Table 3. The model achieved an overall classification accuracy of 84.2%, broken down into classification errors between specific pairs of states via the confusion matrix. The majority of errors (10 days)

⁴ The partitioning of the training and test set is not straightforward, partially caused by the relatively low amount of labeled data. To robustly learn the parameters of the proposed model, the training set has to be temporally contiguous (to train the transition matrix of the HMM) and have enough days for each state to have sufficient training samples for the SVMs of the four different ITCZ states. Only a few of such train/test splits exist, and these are shifted only by a few days from each other and consequently highly correlated. These different train/test set partitions with similar state distributions have little influence on the model's performance (varying not more than +/- 1.5%).

Table 3
Confusion matrix of the VS model on the test set.

–		Expert I			
		Double	Northern	Southern	None
Model	Double	20	2	3	2
	Northern	10	71	0	1
	Southern	0	0	8	0
	None	0	1	0	2
Classification accuracy		101 of 120 (84.2%)			

occurred when the model predicted an nITCZ and the expert labeled it as a dITCZ. The model's classification accuracy of 84.2% is comparable with that of human experts among each other (Table 2). That the model accuracy is slightly higher than that of the best expert (the best human accuracy was 80.7%) could be explained by the fact that the model was trained on labels from Expert I, and thus, the model more closely mimics the labeling behavior of Expert I compared to the other experts who may be using slightly different heuristics for image labeling.

It is also informative to compare the full VS model (using the HsMM/SVM approach, with all features) to simpler variants. Table 4 shows the classification accuracies for various models trained on Expert I's labels, using the same training and test setup as in Table 3. The closest-performing model is an SVM-only model, which classifies each day separately without any temporal (HsMM) component – it is 5% less accurate overall than the HsMM/SVM approach, indicating that including temporal information in the classification approach leads to higher accuracy. Two approaches that each only use a single feature achieve accuracies of 70% and 73%, indicating that the full set of features provide an improvement of 11 to 14% in accuracy. Finally, the standard HMM approach with a multivariate Gaussian for modeling observations given states (using a full covariance matrix) is about 12% less accurate than the HsMM/SVM method.

We focus next on the performance of IR-based models, using the same training and test data from Expert I as used for the VS model. It is important to keep in mind that Expert I considered VS and TPW images as well as IR images when determining the ITCZ state but the IR model only considers IR data. This means that the expert can identify regions of shallow convection while the model cannot. Therefore we expect a decrease in accuracy when the IR model is compared to Expert I's ground truth, particularly when there are weak ITCZ cases present that are visible to the human eye in VS and TPW, but that are not visible in IR imagery alone. The best performance was obtained with a semi-supervised IR model – its performance is 75.8% in terms of classification accuracy on the daily time-scale (Table 5) – this is 8.4% lower than the accuracy of 84.2% obtained by VS model. It is reasonable to assume that the IR model would be more accurate if it were compared to a human expert who identified ITCZ states based only on IR images (see also Table 8). Despite this lower accuracy, given that the IR data records are available for a much longer time-span than VS data, and have fewer data quality issues, the use of the IR model is worth exploring. Furthermore, results from the IR model provide times when a deep convective ITCZ is present, adding additional useful information to the analysis.

Table 4
Classification accuracy of different models.

HsMM/SVM as in paper	84.2%
SVM only	79.2%
HsMM/SVM only with the east–west extent feature for the longest object in the northern and southern image (see appendix)	73.3%
Standard HMM	72.5%
HsMM/SVM only with the average of 12.5° N to 2° N and 2° S to 12.5° S feature (see appendix)	70.0%

Table 5
Confusion matrix of the IR model on the test set.

–		Expert I			
		Double	Northern	Southern	None
Model	Double	9	0	1	1
	Northern	14	71	1	1
	Southern	2	0	8	0
	None	5	3	1	3
Classification accuracy		91 of 120 (75.8%)			

Table 6 shows how the semi-supervised HsMM/SVM IR model performed relative to other variants. As with the VS model, these other approaches have lower accuracy. In particular, the supervised HsMM/SVM approach (which does not take advantage of 7 of the 8 IR images available each day) is 5% less accurate than the semi-supervised approach, indicating that the addition of the unlabeled IR images leads to a more accurate classifier. We conjecture that the reason the SVM only model is more accurate than the version with an HMM is that the HMM propagates some of the incorrect label decisions of the SVM to neighboring states (especially for rare non-present case where experts labeled an ITCZ presence but IR-only features indicate non-presence due to shallow ITCZ occurrences), increasing the error rather than reducing it.

A useful feature of the semi-supervised approach is its ability to incorporate partially labeled data at prediction time (on the test data). For example, the IR model could be used as part of a semi-automated tool whereby a human user labels every k th image and the model infers the rest of the labels. To illustrate this concept, we trained both a semi-supervised HsMM/SVM model and an SVM-only model on fully labeled data (at the daily scale), and then made predictions with each on partially labeled data (again at the daily scale). At prediction time the models were provided with the labels on a subset of images, and then predicted the labels for the rest. The predictions on the rest of the images were then compared to the actual human expert labels from Expert I for these images. The resulting accuracies on the predicted test images are shown in Table 7. The addition of partially labeled data significantly improves the accuracy of the HsMM/SVM approach, from 75.8% with fully unlabeled test data, to 80% with every 2nd day labeled – this increase is because the HsMM component of the model can “propagate” information from the labeled to the unlabeled images at prediction time. In contrast, the supervised SVM-only method cannot leverage these temporal dependencies, and its prediction accuracies remain largely the same (the variation across the columns for SVMs in Table 7 is due only to sampling noise in the training data, i.e., due to predictions being made on different subsets of unlabeled images in each case).

As a final test of our methodology, we applied the supervised VS model and the semi-supervised IR model, both trained on Expert I's labels, to a completely unlabeled image sequence from 2001. We then provided the predictions from each model to Expert I for interpretation. The expert looked at the images using all three image fields (VS, IR, and TPW) and identified which classification decisions from each model were correct or incorrect in the context of the expert's visual inspection. The expert then repeated this analysis, but this time only using the IR images for visual reference, and identified which classifications from the IR model were correct or incorrect relative to visual inspection of the IR images. (We did not ask the expert to examine the VS model prediction relative to IR-only images).

Table 6
Classification accuracy of different models.

Semi-supervised HsMM/SVM as in paper	75.8%
SVM only	72.5%
Supervised HsMM/SVM	70.8%
Standard HMM	65.8%

Table 7
Performance improvement by including expert labels for classification (Accuracy on unlabeled days).

–	No expert labels	Expert labels every 8th day	Expert labels every 2nd day
Semi-supervised HsMM/SVM	75.8%	76.0%	80.0%
SVM only	72.5%	71.2%	73.3%

Table 8 shows the results. On this test, each of the models (VS and IR) was able to achieve 86.7% accuracy relative to Expert I's inspection of the same data (VS or IR). These accuracies may be slightly optimistically biased since the human labeler was shown both the model predictions and the raw images at the same time (whereas previously the labeling was done independently of any model predictions), allowing for the possibility that the models' predictions could bias the human's decisions. Nonetheless the general trend is clear, namely that both VS and IR models can achieve relatively high accuracy (86.7%) when compared to a human looking at the same data that the algorithm has available. When comparing the expert's labels that include the additional information of VS and TPW data with the IR model (IR-only information) less ITCZ cases (11.7%) are detected because of shallow ITCZ cases that show no ITCZ signal in IR images.

3.2. Using the model to study dITCZ climatology over five active seasons

For the purposes of this discussion we focus on the classification labels produced by the VS model over a relatively short illustrative period of 5 seasons (2000–2004). The statistics calculated here are mainly intended to demonstrate the potential of the model rather than an in-depth climatological analysis.

Fig. 4 shows the distribution of the occurrences of the ITCZ states from the model's predictions during the 2000–2004 time period. Red indicates dITCZ days, blue indicates nITCZ days, green indicates sITCZ days, and yellow indicates days with no ITCZ signal. Seasonal totals are shown in the last column and monthly totals (over all 5 years) are shown in the bottom row. Overall, the nITCZ is dominant during this period, accounting for 64% of the days. This is followed by dITCZ (27%), sITCZ (6%) and non-present days (3%).

The bottom row in Fig. 4 provides a general idea of the seasonal evolution of the east Pacific ITCZ through the boreal spring. In February, the dITCZ is present approximately 26% of the time, but this month is dominated by the nITCZ. By March, the occurrence of a dITCZ has increased to 49% and it is now more dominant than the nITCZ. The sITCZ is more present in March than February. The number of dITCZs decreases in April and, by May the ITCZ is located almost exclusively in the north (nITCZ). Based on these five seasons, the peak in dITCZ occurrence happens in March. When considered individually, all years follow this pattern except 2003. In this year the highest number of dITCZs occurs in February and this was the only year where a dITCZ was detected in May. In future studies it would be interesting to see if this general picture holds true or if there are more seasons similar to 2003. It would also be of interest to extend the analysis to January to see if any cases of dITCZ are detected earlier than February.

The last column in Fig. 4 gives an indication of interannual variability. The 2000 season is an outlier compared to the other years. While the number of dITCZ is similar to that in other years (especially 2001 and 2004) the number of nITCZs is greatly reduced and the number of sITCZs is greatly increased. This is the only year in which

Table 8
Model classification accuracy for season 2001 as determined by Expert I.

–	VS model	IR model
Expert I; decision based on all data (VS, IR, TPW)	86.7%	75.0%
Expert I; decision based on IR-only	–	86.7%

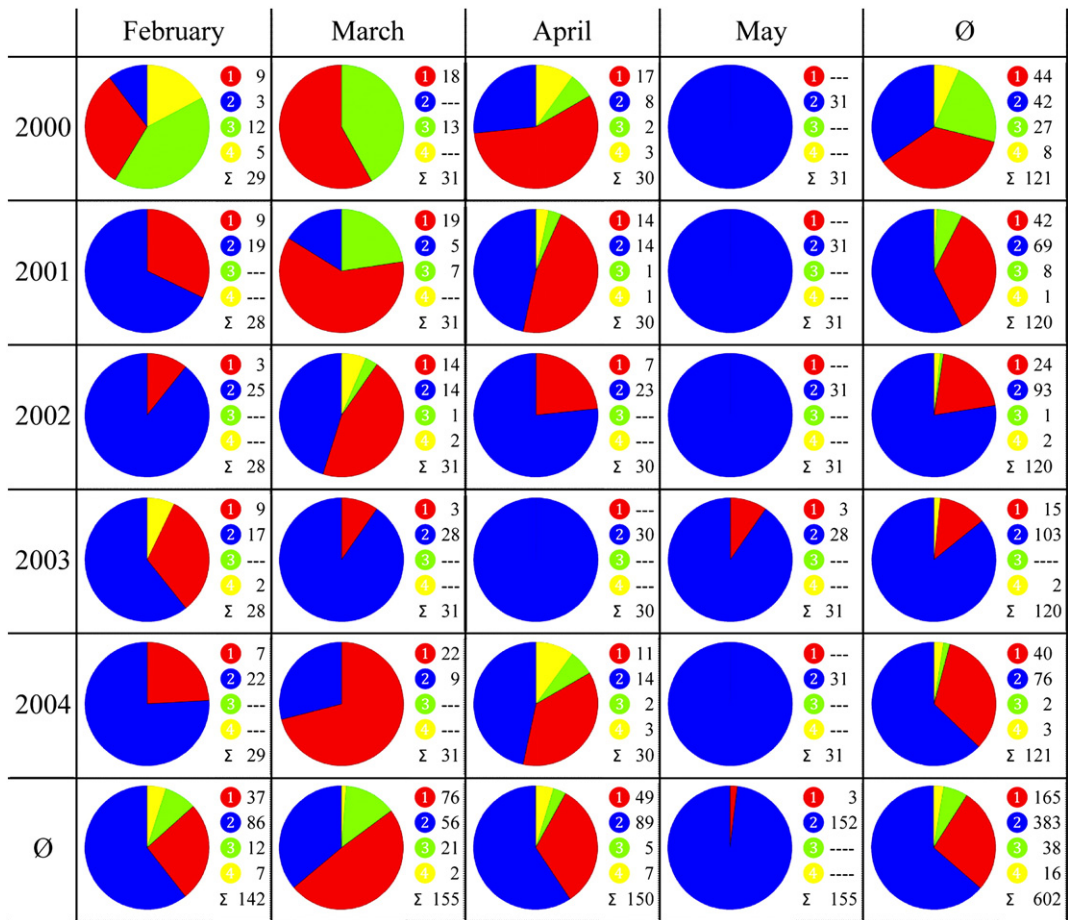


Fig. 4. Intra- and inter-annual occurrences of ITCZ states (red ●: dITCZ; blue ●: nITCZ; green ●: sITCZ; yellow ●: not present). Values on the right of each chart indicate number of days for each ITCZ state.

there are more dITCZs than nITCZs. The 2000 season also has the largest number of days when the ITCZ was labeled ‘not present’. The 2003 season has the fewest dITCZs and sITCZs out of all five seasons.

It is possible that the large number of dITCZs and sITCZs seen in 2000 (and maybe even 2001) are due to a somewhat strong La Niña during the 1999/2000 winter. A longer time series analysis could better

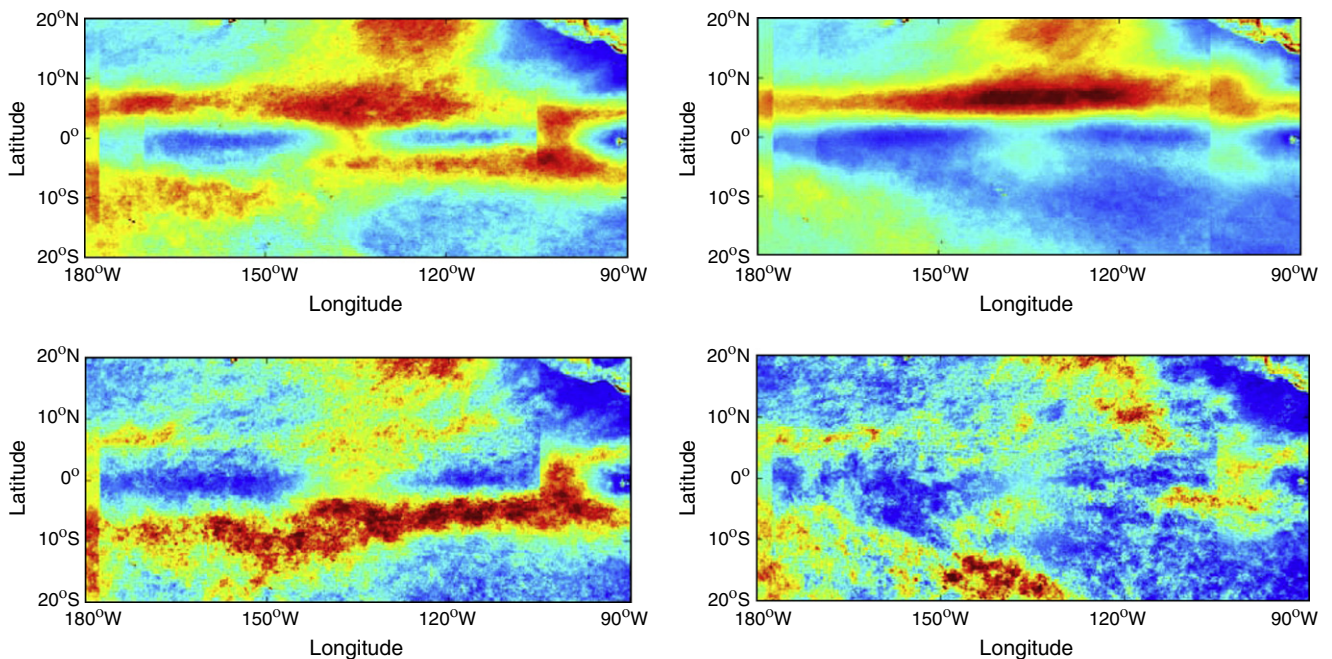


Fig. 5. Average VS images for each dITCZ state: double, top left; north, top right; south, bottom left; no presence, bottom right.

clarify this hypothesis and indicate the effects of El Niño – Southern Oscillation (ENSO) on the different states of the east Pacific ITCZ.

Fig. 5 shows composite VS satellite images for each of the four states, showing the average of the VS intensity. The top left plot shows the dITCZ with two zonally elongated cloud bands on each side of the equator. From this composite plot one can observe a longitudinal offset between the northern and southern parts of the dITCZ such that the northern part appears to be located west of the southern part. The top right panel shows the typical nITCZ. The bottom left panel shows the sITCZ. It is interesting to note that the composite sITCZ shown here looks spatially different from the southern part of the dITCZ in the top left panel. This introduces the question of whether there is asymmetry in ITCZ structure in one hemisphere depending on presence or absence of a simultaneous ITCZ structure in the other hemisphere. Finally the bottom right panel shows the average VS field when no ITCZ is present. The output images are generally consistent with the limited prior climatological knowledge we have for each state, providing additional validation of the model.

In Fig. 6 the average VS images after the morphologic reconstruction step are displayed, filtering out non-ITCZ clouds and emphasizing the ITCZ signal. In both Figs. 5 and 6 some vertical artifacts are visible at the location where measurements from two different GOES satellites are combined. This type of artifact has been cleaned up in the more consistent IR output in the GridSat dataset (Knapp et al., 2011). Suppressing such artifacts is not of direct relevance to the results presented here; for the interested reader, additional information on this topic can be found in Minnis (1989) and Govaerts et al. (2008).

Furthermore, we briefly demonstrate the benefits of the individual features discussed in Section 2 by examining the geometry of the dITCZ, using the northern and southern centers of mass of dITCZ structures. Specifically, given the ITCZ states as classified by the model, we investigate if the shift between the northern and southern dITCZ, human experts observed while labeling the data (and rudimentary visible in the composite images of Figs. 5 and 6, top left), also appears in the feature statistics. Out of the 165 dITCZ days during the five seasons, in 109 cases (66%) the center of mass of the southern dITCZ was located to the east of the northern dITCZ (we define this as a ‘positive offset’) and the two centers were, on average, 23.4° apart in longitude. The other 56 cases (34%) have the southern part of dITCZ

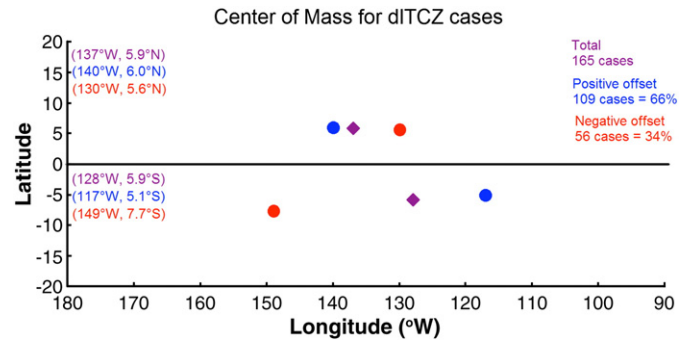


Fig. 7. dITCZ cases for Feb–May, 2000–2004. Mean location of the centers of mass of the north part and south part of the dITCZ is depicted for all cases (purple), for positive offset cases (blue, when north part is west of south part) and for negative offset cases (red, when north part is east of south part).

located to the west of the northern part with an 18.9° mean longitudinal difference in their centers of mass (‘negative offset’). Fig. 7 illustrates the mean center of mass in each hemisphere for all of the dITCZ cases and the positive and negative offset cases. Note that the mean location of the northern dITCZ does not change drastically between the positive and negative offset cases (change in longitude: 10°). It is the southern dITCZ location that shifts widely between positive and negative offset cases (change in longitude: 32°).

4. Conclusion

We have introduced an algorithm to extract the spatial location and expansion of the dITCZ phenomenon and to track its temporal behavior. For spatial segmentation, we developed a fully automatic unsupervised method based on the backbone path method. The extracted ITCZ regions are consistent and robust, and form – with additional derived features – the observation set for the temporal segmentation. The HsMM/SVM hybrid approach performed best in our experiments for this temporal task. We showed that the classification labels from this model are comparable in reliability to those from human experts. A VS model was developed that identifies ITCZ states

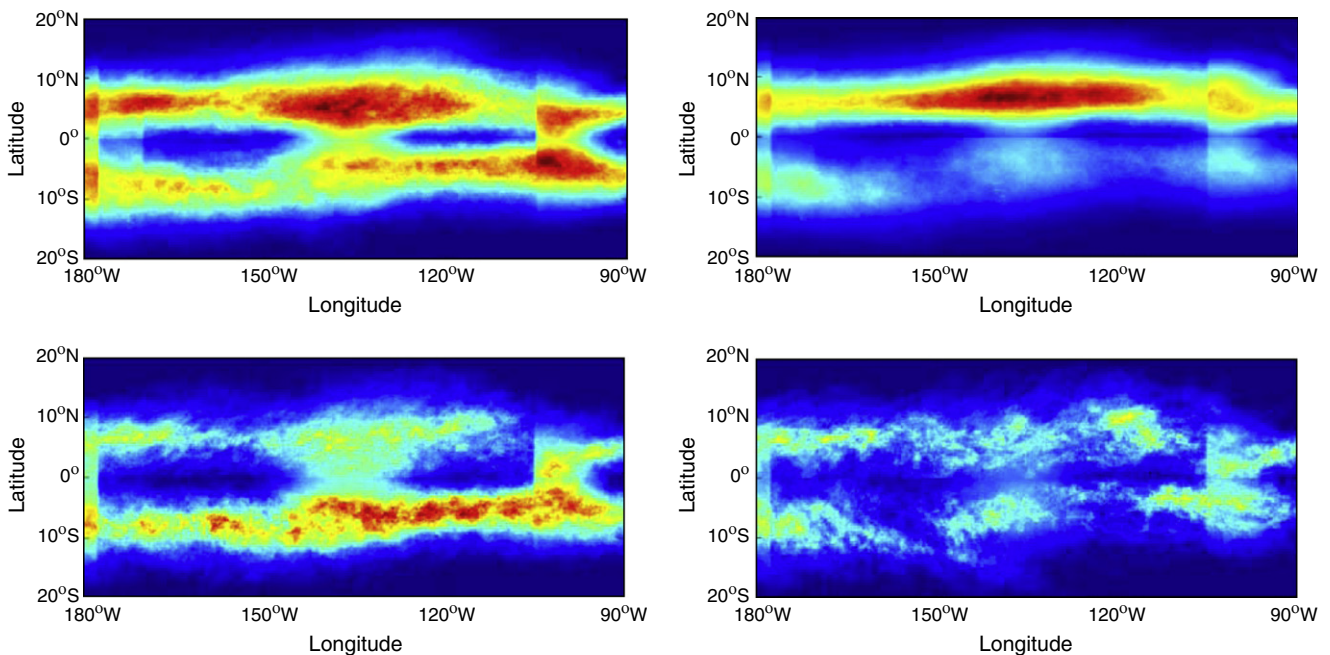


Fig. 6. Average VS morphologic reconstructed images for each dITCZ state: double, top left; north, top right; south, bottom left; no presence, bottom right.

even when there is only shallow convection present. A corresponding IR model performs the same task but only identifies the ITCZ states when deep convection is present. Both models proved very accurate (well within the range of human labeling variability) when compared to expert opinion using the appropriate field (either VS or IR).

The model's output for the period of 2000–2004 shows clear evidence that this method would be useful in classifying the east Pacific ITCZ states. The results suggest that it is worth conducting an in-depth long-term analysis of dITCZ phenomena, using both of the models developed here with VS and IR data available (with some gaps) since 1980. Recent studies dealing with the temporal behavior of the dITCZ phenomenon used either (a) a larger temporal time scale (Gu et al., 2005), (b) rather general image statistics (Chen et al., 2008) neglecting the ITCZ signal itself and the temporal dependencies, or (c) manual identification (Wang & Magnusdottir, 2006). In contrast, with the model proposed in this paper, it is now feasible to automatically detect the dITCZ phenomenon and track its signal on a daily time scale, in turn computing meaningful statistics about location, extent and temporal distribution of the dITCZ phenomenon. Furthermore, this type of analysis can help us better understand the mechanisms causing the dITCZ phenomena and thus to improve existing weather and climate models for the tropical circulation system.

Future work may include fine tuning of the model in terms of feature selection, improved methods for classification of rare cases

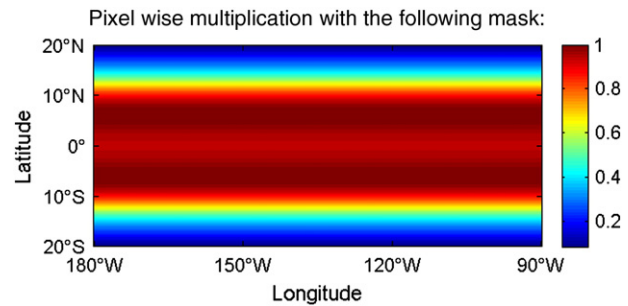
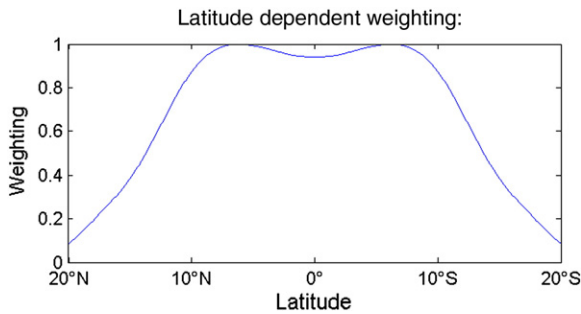
(such as 'no ITCZ'), as well as a detailed comparison of the unsupervised backbone method developed in this work with the Markov random field in Bain et al. (2011). The results achieved with the proposed HMM-based method for detecting and characterizing the ITCZ signals on both hemispheres encourage an in-depth analysis to investigate if initial hypotheses derived from the five season analysis can be proven on long-term data. Initial hypotheses include the accumulation of sITCZ events in La Niña years and the asymmetry in spatial structure of the dITCZ depending on relative longitudinal offset across the equator. With a longer time series we will be able to study the effects of ENSO in general on the four different classes of double ITCZ variability. We will examine the effects of propagation of the Madden Julian Oscillation (MJO) on the double ITCZ variability. Furthermore, with the temporally finer resolution IR channel we can examine whether the results of Bain et al. (2010) for the diurnal cycle of the boreal summer ITCZ hold up for the weaker boreal spring convergence zones.

Acknowledgments

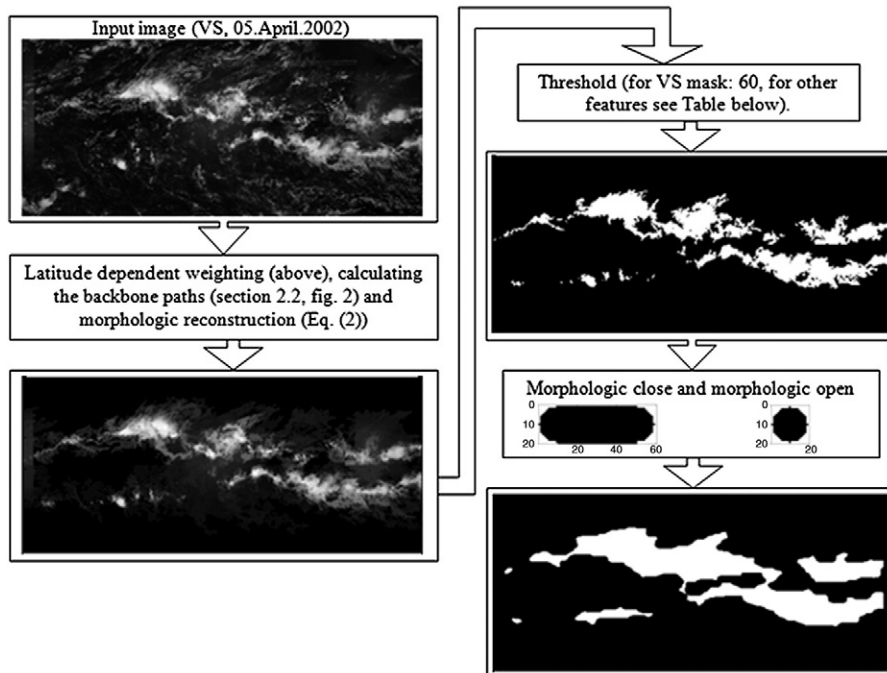
We thank Ashley Payne and Yi-Hui Wang for help with labeling satellite images. We thank Dr. Ken Knapp of NOAA National Climate Data Center for providing the GridSat data. We thank two anonymous reviewers for comments on the manuscript. The project was originally supported by NSF grant ATM-0530926.

Appendix A. Image segmentation

Pre-processing



Processing chain for the feature extraction:



Appendix B. Feature calculation

Details on the calculation of selected features:

- **Border tracing** (Klette & Rosenfeld, 2004)

Let (p,r) be the first edge where a 0/1 switch occurs:

$$B_t(p) = 0 \rightarrow B_t(r) = 1 \tag{A.1.1}$$

Let $\xi(p)$ be the 4-neighborhood of p in circular order:

$$\xi(p) = \{\rho_0, \rho_1, \rho_2, \rho_3 = r\} \tag{A.1.2}$$

With the 4-neighborhood being defined as follows:

$$\begin{matrix} * & \rho_0 & * \\ \rho_3 & p & \rho_1 \\ * & \rho_2 & * \end{matrix}$$

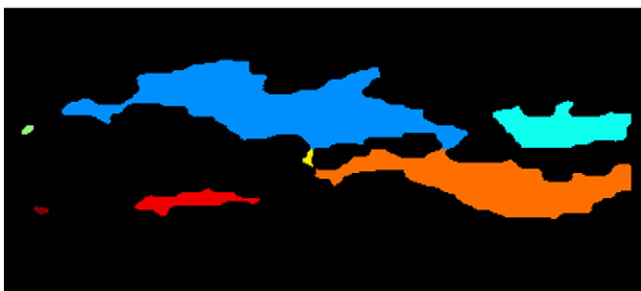
Find the first $r^* := \rho_k$ in $\xi(p)$ where $\rho_k = 1$ holds : $\min_{0 \leq k \leq 3} \rho_k = 1$ and add r^* to the list of border points. (A.1.3)

If r^* is equal to the initial point of Eq. (A.1.1) the complete object border was traced and the next 0/1 switch in the image needs to be found to extract the next object starting from Eq. (A.1.1). Otherwise set $r = p$ and $p = r^*$ and repeat steps (A.1.2) and (A.1.3) until r^* equals the initial point of Eq. (A.1.1)

$$r^* = \text{initial } p? \begin{cases} \text{yes} & \rightarrow \text{algorithm finished, object border traced} \\ \text{no} & \rightarrow r := p, p := r^* \text{ and proceed with Eq. (A.1.2)} \end{cases} \tag{A.1.4}$$

The image is scanned from left-top to right-bottom until all objects are traced.

Result of border tracing:



- **Fourier descriptors**

The extracted object borders are interpreted as closed-form contours of periodic functions $x(t)$ in the complex plane. For a continuous line in the complex space the Fourier coefficients c_n are defined as:

$$c_n = \frac{1}{T} \int_0^T x(t) e^{-2\pi i n t / T} dt \tag{A.2.1}$$

with T being the perimeter of the continuous line

Due to the discrete character of the border line, we obtain a set of nodes $\{x_1, x_2, \dots, x_N = x_0\}$ in the complex plane connected by section-wise

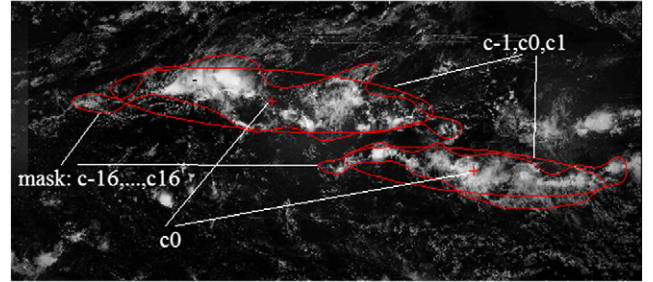
straight lines instead. Thus, an explicit calculation of the section-wise straight lines has to be conducted and it can be shown (e.g. Arbter et al., 1990) that the Fourier coefficients c_n can be efficiently computed by:

$$c_0 = \frac{1}{2T} \sum_{k=0}^{N-1} (x_k + x_{k+1}) |x_{k+1} - x_k|$$

$$c_n = \frac{T}{(2\pi n)^2} \sum_{k=0}^{N-1} \left(\frac{x_k - x_{k-1}}{|x_k - x_{k-1}|} - \frac{x_{k+1} - x_k}{|x_{k+1} - x_k|} \right) e^{-2\pi i n t_k / T} \tag{A.2.2}$$

with $t_k = \sum_{l=0}^{k-1} |x_{l+1} - x_l|$ for $k > 0$, $t_0 = 0$

Fourier descriptors:



- **Moments**

Central and invariant moments can provide useful information. The first order moment corresponds to the center of mass of the image I :

$$\bar{x} = \frac{\sum_x \sum_y x I(x,y)}{\sum_x \sum_y I(x,y)} \text{ and } \bar{y} = \frac{\sum_x \sum_y y I(x,y)}{\sum_x \sum_y I(x,y)} \tag{A.3.1}$$

Higher order moments are less easy to interpret geometrically but nevertheless contain valuable information about the current state of the ITCZ.

$$\mu_{pq} = \sum_x \sum_y (x - \bar{x})^p (y - \bar{y})^q I(x,y) \tag{A.3.2}$$

To achieve invariance to scale and translation the central moments are scaled:

$$\eta_{pq} = \frac{\mu_{pq}}{\mu_{00}^{1+(p+q)/2}} \tag{A.3.3}$$

Furthermore by combining the scale invariant moments a set of rotationally invariant moments – the Hu set of moments (Hu, 1962) – can be computed. A subset of these moments is used for classification in this paper.

List of features used for the models:

The table below defines the lists of features used for each of the VS and IR models used in the paper. The final set of features was selected from an initial set of 177 features based on the following simple step-wise algorithm.

- Build a classifier with all of the features
- For each feature
 - Deselect the feature
 - If the classification on the training set improves exclude the feature
 - Otherwise include the feature in the list again
- Proceed until the last feature in the list is reached

This greedy feature selection method is certainly not optimal, but provided a simple method for feature selection for our problem. More sophisticated feature selection methods could likely be used to further improve the performance of the algorithms described in this paper.

Definition. Complete image: 20° N to 20° S; northern image: 20° N to 0°; southern image: 0° to 20° S

VS model feature list:

ID	Feature description	Trend north vs. not-north	Trend south vs. not-south
<i>(General image statistics)</i>			
1	Average pixel value of 12.5° N to 2° N	↑	↓
2	Average pixel value of 2° S to 12.5° S	↓	↑
3	Latitude of the center of mass of the complete image	↓	↑
4	η_{03} of Eq. (A.3.3) for the complete image	↑	↓
5	Latitude of the center of mass of the northern image	↑	↑
6	η_{02} for the southern image	↑	↓
7	η_{20} for the southern image	↑	↓
<i>(Based on the backbone path)</i>			
8	Regression line parameters of the optimal path (complete image): $\text{linreg}(P^0)$ with P^0 result of Eq. (1.2)	↓	↑
9	Mean of pixel intensities along the optimal path (northern image)	↑	↓
10	Variance of pixel intensities along the optimal path (northern image)	↑	↓
11	Mean of pixel intensities along the optimal path (southern image)	↑	↓
<i>(Based on the extracted binary mask, with two different thresholds: 60 and 100 on the scaled reconstructed image with values in the interval [0 255])</i>			
<i>Definition of "longest object": The object with the largest east–west extent of all extracted objects</i>			
<i>Features 12–17: Northern image with threshold 100</i>			
12	East–west extent of the longest object	↑	↓
13	Area (in pixels) of the longest object	↑	↓
14	Perimeter (in pixels along the border line) of the longest object	↑	↓
15	Number of direction changes in the border line of the longest object	↑	↓
16	Length of the major axis of the ellipse of the longest object derived by the first Fourier descriptors c_0, c_{-1}, c_1	↑	↓
17	Sum of the east–west extent of all objects with an area larger than 500 pixels	↑	↓
<i>Features 18–26: Southern image with threshold 100</i>			
18	Compactness of the longest object	↑	↓
19	Perimeter of the longest object	↓	↑
20	Number of direction changes in the border line of the longest object	↓	↑
21	Length of the major axis of the ellipse of the longest object derived by the first Fourier descriptors c_0, c_{-1}, c_1	↓	↑
22	Ratio of major/minor axis of the ellipse of the longest object (Fourier descriptor)	↓	↑
23	Major axis' inclination of the ellipse of the longest object (Fourier descriptor)	↓	↑
24	Number of objects with an area larger than 500 pixels	↑	↓
25	Sum of the perimeter of all objects with an area larger than 500 pixels	↓	↑
26	Sum of the east–west extent of all objects with an area larger than 500 pixels	↓	↑
<i>Features 27–32: Northern image with threshold 60</i>			
27	Area of the longest object	↑	↓
28	Compactness of the longest object	↓	↑
29	Perimeter of the longest object	↑	↓
30	Sum of the area of all objects with an area larger than 500 pixels	↑	↓
31	Sum of the perimeter of all objects with an area larger than 500 pixels	↑	↓
32	Sum of the east–west extent of all objects with an area larger than 500 pixels	↑	↓
<i>Features 33–34: Southern image with threshold 60</i>			
33	Sum of the area of all objects with an area larger than 500 pixels	↓	↑
34	Sum of the east–west extent of all objects with an area larger than 500 pixels	↓	↑

IR model feature list:

ID	Feature description	Trend north vs. not-north	Trend south vs. not-south
<i>(General image statistics)</i>			
1	Average pixel value of 2° S to 12.5° S	↓	↑
2	Latitude of the center of mass of the complete image	↓	↑
3	η_{02} of Eq. (A.3.3) for the complete image	↓	↑
4	Latitude of the center of mass of the northern image	↑	↓
5	η_{12} for the northern image	↓	↑
6	η_{03} for the northern image	↓	↑
7	Hu moment $J_1 = \eta_{02} + \eta_{20}$ (northern image)	↓	↑
8	Hu moment $J_2 = (\eta_{20} - \eta_{02})^2 + (2\eta_{11})^2$ (northern image)	↓	↑
9	Hu moment $J_4 = (\eta_{30} + \eta_{12})^2 + (\eta_{21} + \eta_{03})^2$ (northern image)	↑	↓
10	Latitude of the center of mass of the southern image	↑	↓
11	η_{02} for the southern image	↑	↓
12	η_{20} for the southern image	↑	↓
13	Hu moment $J_1 = \eta_{02} + \eta_{20}$ (southern image)	↑	↓
<i>(Based on the backbone path)</i>			
14	Mean of pixel intensities along the optimal path (northern image)	↑	↓
15	Variance of pixel intensities along the optimal path (northern image)	↑	↓

Appendix B (continued)

ID	Feature description	Trend north vs. not-north	Trend south vs. not-south
<i>(Based on the extracted binary mask, with two different thresholds: 30 and 18 on the scaled reconstructed inverted IR image with values in the interval [0 255])</i>			
<i>Definition of "longest object": The object with the largest east–west extent of all extracted objects</i>			
<i>Features 16–20: Northern image with threshold 30</i>			
16	East–west extent of the longest object	↑	↓
17	Area (in pixels) of the longest object	↑	↓
18	Perimeter (in pixels along the border line) of the longest object	↑	↓
19	Major axis' inclination of the ellipse of the longest object (Fourier descriptor)	↑	↓
20	Sum of the area of all objects with an area larger than 500 pixels	↑	↓
<i>Features 21–28: Southern image with threshold 30</i>			
21	Compactness of the longest object	↑	↓
22	Perimeter of the longest object	↓	↑
23	Number of direction changes in the border line of the longest object	↓	↑
24	Length of the major axis of the ellipse of the longest object derived by the first Fourier descriptors c_0, c_{-1}, c_1	↓	↑
25	Major axis' inclination of the ellipse of the longest object (Fourier descriptor)	↓	↓
26	Sum of the area of all objects with an area larger than 500 pixels	↓	↑
27	Sum of the perimeter of all objects with an area larger than 500 pixels	↓	↑
28	Sum of the east–west extent of all objects with an area larger than 500 pixels	↓	↑
<i>Features 29–34: Northern image with threshold 18</i>			
29	Perimeter of the longest object	↑	↓
30	Length of the major axis of the ellipse of the longest object derived by the first Fourier descriptors c_0, c_{-1}, c_1	↑	↓
31	Major axis' inclination of the ellipse of the longest object (Fourier descriptor)	↑	↓
32	Sum of the perimeter of all objects with an area larger than 500 pixels	↑	↓
33	Sum of the east–west extent of all objects with an area larger than 500 pixels	↑	↓
34	Ratio of the sum of the east–west to the north–south extent of all objects with an area larger than 500 pixels	↑	↓
<i>Features 35–37: Southern image with threshold 18</i>			
35	East–west extent of the longest object	↓	↑
36	Sum of the area of all objects with an area larger than 500 pixels	↓	↑
37	Sum of the east–west extent of all objects with an area larger than 500 pixels	↓	↑

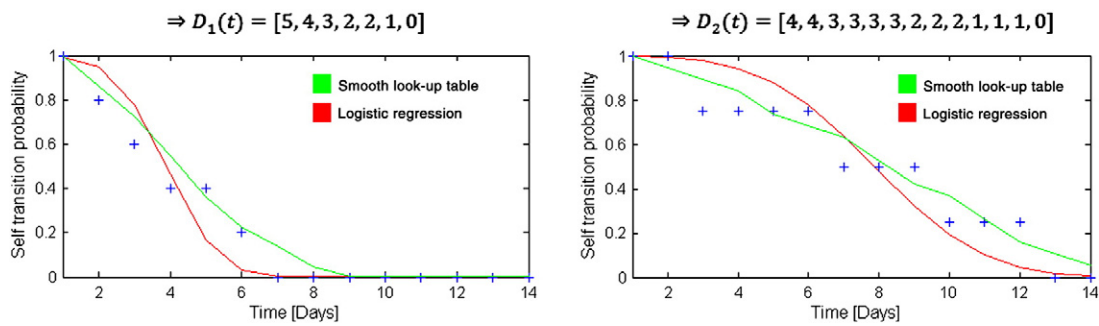
Appendix C. HsMM details

Description of the models:

Self-transition duration estimate (simple example, 2 states only, 46 labels):

Top: training sequence L , Bottom: $d_k(i)$'s

1 1 1 1 1	2 2	1 1 1	2 2 2 2 2 2 2 2	1 1	2 2 2 2 2 2	1	2 2 2 2 2 2 2 2 2 2 2 2	1 1 1 1 1 1
$d_1(1)=5$	$d_1(2)=2$	$d_2(1)=3$	$d_2(2)=9$	$d_3(1)=2$	$d_3(2)=6$	$d_4(1)=1$	$d_4(2)=12$	$d_5(1)=6$



For the example above the differences between the fitted logistic regression and the look-up table are rather small since there are no outliers in the example – for the real data with few training examples, the look-up table provides robustness against outliers.

VS model parameters used in the experiments:

$\pi = [0.3058 \quad 0.4959 \quad 0.1405 \quad 0.0579]$

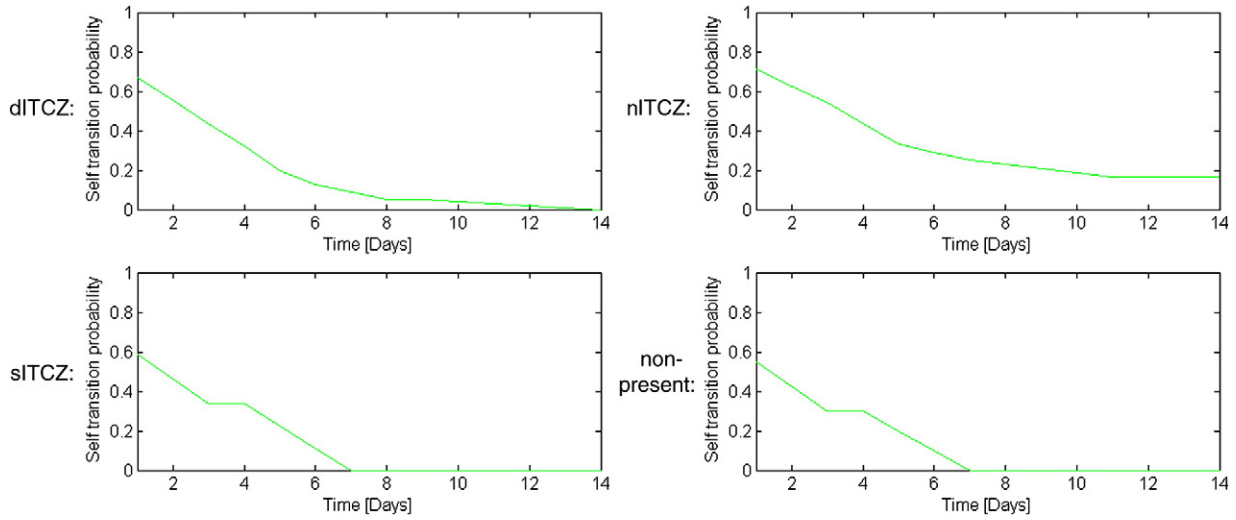
$A_{static} = \begin{bmatrix} 0.6389 & 0.0833 & 0.2222 & 0.0556 \\ 0.0833 & 0.9000 & 0.0000 & 0.0167 \\ 0.4706 & 0.0000 & 0.4118 & 0.1176 \\ 0.1429 & 0.4286 & 0.1429 & 0.2857 \end{bmatrix}; \quad \lambda = 0.1$

• Emission probabilities by SVMs (parameters for each state are learned by the toolbox by De Brabanter et al., 2010)

• Logistic regression parameters of Eq. (6.4):

$\epsilon_{DIRCZ} = 10.48$	$\epsilon_{north} = 4.84$	$\epsilon_{south} = 9.85$	$\epsilon_{none} = 1.49$
$\gamma_{DIRCZ} = -1.31$	$\gamma_{north} = -0.04$	$\gamma_{south} = -1.66$	$\gamma_{none} = 1.51$

- Self-transition look-up tables:



IR model parameters used in the experiments:

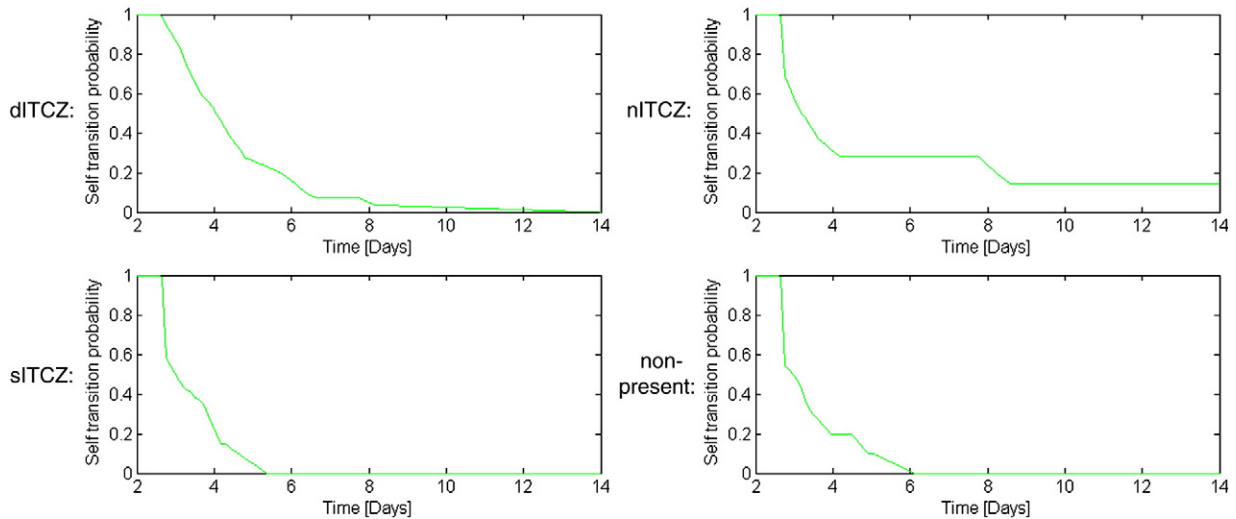
- $\pi = [0.3337 \quad 0.4898 \quad 0.1163 \quad 0.0602]$

- $A_{static} = \begin{bmatrix} 0.9572 & 0.0122 & 0.0245 & 0.0061 \\ 0.0104 & 0.9875 & 0.0000 & 0.0021 \\ 0.0702 & 0.0000 & 0.9123 & 0.0175 \\ 0.0169 & 0.0508 & 0.0169 & 0.9153 \end{bmatrix}; \quad \lambda = 1$

- Emission probabilities by SVMs (parameters for each state are learned by the toolbox by De Brabanter et al., 2010)
- Logistic regression parameters of Eq. (6.4):

$\varepsilon_{dITCZ} = 5.92$	$\varepsilon_{north} = 6.45$	$\varepsilon_{south} = 4.92$	$\varepsilon_{none} = 1.03$
$\gamma_{dITCZ} = -0.22$	$\gamma_{north} = -0.46$	$\gamma_{south} = -0.56$	$\gamma_{none} = 1.96$

- Self-transition look-up tables:



Appendix D. Detailed description of the semi-supervised EM algorithm

E-Step (Find \mathbf{Q}^r given \mathbf{O} and θ^r under the constraint: \mathbf{Q}^r traverses \mathbf{L}^{known})

- Calculate the observation probabilities $b_i(\mathbf{x}) = P(\mathbf{o}_t = \mathbf{x} | q = i)$ for all time steps t and all states i using the SVMs.
- Force the state sequence \mathbf{Q} to traverse the supervised labels \mathbf{L}^{known} , by setting the transition probabilities at given labeling times to 1 for the transition into the labeled state and to 0 otherwise:

$$\forall j, k: a_{j, l_{8^*k+m}} = P(q_{8^*k+m} = l_{8^*k+m} | q_{8^*k+m-1} = j) = 1 \quad (\text{A.4.1})$$

$$a_{j, other} = P(q_{8^*k+m} \neq l_{8^*k+m} | q_{8^*k+m-1} = j) = 0$$

- By doing so, the probabilities $P(\mathbf{O}, \mathbf{Q} | \theta)$ are 0 for all sequences \mathbf{Q} not fulfilling the supervised label constraint. Another interpretation is that for our likelihood maximization problem we actually marginalize only over the possible state sequences $\mathbf{Q}^{supervised}$, i.e.:

$$\forall \mathbf{Q} \neq \mathbf{Q}^{supervised}: P(\mathbf{O}, \mathbf{Q} | \theta) = 0 \Rightarrow \mathcal{L}(\theta; \mathbf{O}) = \sum_{\forall \mathbf{Q}} P(\mathbf{O}, \mathbf{Q} | \theta) \quad (\text{A.4.2})$$

$$= \sum_{\forall \mathbf{Q}^{supervised}} P(\mathbf{O}, \mathbf{Q} | \theta)$$

- Now either the alpha-beta or the Viterbi algorithm can be used to estimate the most probable state sequence \mathbf{Q} which satisfies the condition of traversing all supervised labels \mathbf{L}^{known} . Since for the Viterbi algorithm the constraint that a minimum of 8 consecutive steps in the same state can be guaranteed, in our application we use it instead of the alpha-beta algorithm (which does not enforce this constraint) and we can obtain directly from the state sequence \mathbf{Q} a new training sequence $\mathbf{L}^{r+1} = \mathbf{Q}$ for the M-Step.

M-step:

- Update the model parameters θ^{r+1} as follows:
 - Train the SVM parameters \mathbf{w}_i and u_i by using the labels \mathbf{L}^{r+1} from the E-Step and the corresponding complete observation set \mathbf{O} . Set up the training pairs and learn the SVMs as for the VS case (Eqs. 6.1–6.4).
 - Learn \mathbf{A} and $\boldsymbol{\pi}$ using the labels \mathbf{L}^{r+1} equivalent to the supervised VS case (Eqs. 4.1–4.2).
 - Force the minimum self-transition time to be at least 8 time steps: $\forall i, 0 < t < 8: D_i(t) = 1$

References

- Arbter, K., Snyder, W. E., Burkhardt, H., & Hirzinger, G. (1990). Application of affine-invariant Fourier descriptors to recognition of 3-D objects. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 12, 640–647.
- Bain, C. L., De Paz, J., Kramer, J., Magnusdottir, G., Smyth, P., Stern, H., et al. (2011). Detecting the ITCZ in instantaneous satellite data using spatiotemporal statistical modeling: ITCZ climatology in the east Pacific. *Journal of Climate*, 24, 216–230.
- Bain, C., Magnusdottir, G., Smyth, P., & Stern, H. (2010). Diurnal cycle of the Intertropical Convergence Zone in the east Pacific. *Journal of Geophysical Research*, 115, 1–10.
- Bourlard, H. A., & Morgan, N. (1994). *Connectionist speech recognition: A hybrid approach*. Kluwer Academic Publishers.
- Chen, B., Lin, X., & Bacmeister, J. T. (2008). Frequency distribution of daily ITCZ patterns over the western-central Pacific. *Journal of Climate*, 21, 4207–4222.
- De Brabanter, K., Karsmakers, P., Ojeda, F., Alzate, C., De Brabanter, J., Pelckmans, K., et al. (2010). *LS-SVMlab toolbox user's guide version 1.7*. Internal Report 10-146, ESAT-SISTA. Belgium: K.U.Leuven. <http://www.esat.kuleuven.be/sista/lssvmlab/> (04.02.2011).
- Govaerts, Y. M., Lattanzio, A., Taberner, M., & Pinty, B. (2008). Generating global surface albedo products from multiple geostationary satellites. *Remote Sensing of Environment*, 112(6), 2804–2816.
- Gu, G., Adler, R. F., & Sobel, A. H. (2005). The eastern Pacific ITCZ during the boreal spring. *Journal of Atmospheric Sciences*, 62, 1157–1174.
- Hu, M. K. (1962). Visual pattern recognition by moment invariants. *IRE Transactions on Information Theory*, 8, 179–187.
- Hubert, L. F., Krueger, A. F., & Winston, J. S. (1969). The double Intertropical Convergence Zone – Fact or fiction? *Journal of Atmospheric Sciences*, 26, 771–773.
- Klette, R., & Rosenfeld, A. (2004). *Digital geometry: Geometric methods for digital picture analysis*. Morgan Kaufmann.
- Knapp, K. R., Ansari, S., Bain, C. L., Bourassa, M. A., Dickinson, M. J., Funk, C., et al. (2011). Globally gridded satellite observations for climate studies. *Bulletin of the American Meteorological Society*, 92, 893–907.
- Lietzke, C. E., Deser, C., & Vonder Haar, T. H. (2001). Evolutionary structure of the eastern Pacific double ITCZ based on satellite moisture profile retrievals. *Journal of Climate*, 14, 743–751.
- Liu, W. T., & Xie, X. (2002). Double Intertropical Convergence Zones a new look using scatterometer. *Geophysical Research Letters*, 29, 2072.
- Magnusdottir, G., & Wang, C. -C. (2008). Intertropical Convergence Zones during the active season in daily data. *Journal of Atmospheric Sciences*, 65, 2425–2436.
- Masunaga, H., & L'Ecuyer, T. S. (2010). The southeast Pacific warm band and double ITCZ. *Journal of Climate*, 23, 1189–1208.
- Minnis, P. (1989). Viewing zenith angle dependence of cloudiness determined from coincident GOES East and GOES West data. *Journal of Geophysical Research*, 94, 2303–2320.
- Murphy, K. P. (2002). Hidden semi-Markov models (HSMs). <http://www.ai.mit.edu/murphyk> (04.02.2011).
- Platt, J. (1999). Probabilistic outputs for support vector machines and comparisons to regularized likelihood methods. In A. J. Smola, P. Bartlett, B. Schölkopf, & D. Schuurmans (Eds.), *Advances in Large Margin Classifiers*. MIT Press.
- Rabiner, L. R. (1989). A tutorial on hidden Markov models and selected applications in speech recognition. *Proceedings of the IEEE*, 77, 257–286.
- Robinson, K., & Whelan, P. F. (2004). Efficient morphological reconstruction: A downhill filter. *Pattern Recognition Letters*, 25, 1759–1767.
- Wang, C. -C., & Magnusdottir, G. (2006). The ITCZ in the central and eastern Pacific on synoptic time scales. *Monthly Weather Review*, 134, 1405–1421.
- Zhang, C. (2001). Double ITCZs. *Journal of Geophysical Research*, 106, 11785–11792.
- Zhong, S. (2005). Semi-supervised sequence classification with HMMs. *International Journal of Pattern Recognition and Artificial Intelligence*, 19, 165–182.