The Development and Assessment of a Model-, Grid-, and Basin-Independent Tropical Cyclone Detection Scheme

K. J. Tory, S. S. Chand, R. A. Dare, and J. L. McBride

Centre for Australian Weather and Climate Research, Bureau of Meteorology, Melbourne, Victoria, Australia

(Manuscript received 30 July 2012, in final form 10 December 2012)

ABSTRACT

A novel approach to tropical cyclone (TC) detection in coarse-resolution numerical model data is introduced and assessed. This approach differs from traditional detectors in two main ways. First, it was developed and tuned using 20 yr of ECMWF Interim Re-Analysis (ERA-Interim) data, rather than using climate model data. This ensures that the detector is independent of any climate models to which it will later be applied. Second, only relatively large-scale parameters resolvable in climate models are included, in order to minimize any grid-resolution dependence on parameter thresholds. This approach is taken in an attempt to construct a unified TC detection procedure applicable to all climate models without the need for any further tuning or adjustment.

Unlike traditional detectors that seek to identify TCs directly, the authors’ method seeks to identify conditions favorable for TC formation. Favorable TC formation regions at the center of closed circulations in the lower troposphere to the midtroposphere are identified using a low-deformation vorticity parameter. Additional relative and specific humidity thresholds are applied to ensure the thermodynamic environment is favorable, and a vertical wind shear threshold is applied to eliminate storms in a destructive shear environment. A further requirement is that thresholds for all parameters must be satisfied for at least 48 h before a TC is deemed to have developed.

A thorough assessment of the detector performance is provided. It is demonstrated that the method reproduces realistic TC genesis frequency and spatial distributions in the ERA-Interim data. Application of the detector to four climate models is presented in a companion paper.

1. Introduction

The possibility of an increased threat from tropical cyclones (TCs) in a warming climate has fuelled considerable interest in potential changes in future TC behavior. TC projection techniques rely on global climate models to provide a best estimate of the future state of the climate. While these models are capable of producing TC-like circulations (e.g., Manabe et al. 1970; Bengtsson et al. 1982, 1995, 2007; Krishnamurti 1988; Krishnamurti et al. 1989; Broccoli and Manabe 1990; Wu and Lau 1992; Haarsma et al. 1993; Tsutsui and Kasahara 1996; Vitart et al. 1997, 1999; Sugi et al. 2002; Camargo and Zebiak 2002), the scale and fine structure of the model TC is compromised by the relatively coarse resolution, which makes them difficult to identify.

Sophisticated TC detection schemes are required to identify TC-like characteristics in coarse-resolution models and to judge whether the circulation should be considered a TC. It is difficult to assess the performance of TC detectors that have been tuned in climate model data to best reproduce observed TC climatology, as there is no way to separate model error from detection error. This error uncertainty adds to the uncertainty of the TC projections. However, the projection uncertainty can be reduced by analyzing multiple models [e.g., the model intercomparison project of Walsh et al. (2012)], and it might be reduced further by using an alternative TC detector that differs fundamentally in design. In this paper, we develop and tune such a detector proposed by Tory et al. (2013b, hereafter T13b), but first we provide background information on more traditional TC detection techniques.

a. Traditional TC detectors

Most recent TC detectors follow the method of Bengtsson et al. (1995), which tests for TC characteristics
such as enhanced relative vorticity in the lower troposphere, a local minimum in surface pressure near the storm center, wind or temperature anomalies consistent with a warm-cored structure, and some minimum lifetime limit. In most TC detectors at least one grid-space-dependent threshold (e.g., surface wind speed and/or lower troposphere relative vorticity) is required to take into account the limit to which models of varying grid spacing can resolve the TC circulation (e.g., Chauvin et al. 2006; Walsh et al. 2007; Zhao et al. 2009; Murakami and Wang 2010; Murakami et al. 2011).

The tuning of grid-space-dependent thresholds is typically achieved by adjustment to best match the model detected TC climatology with observed TC climatology. However, tuning within the GCM introduces two inseparable errors: the TC definition error and the GCM modeled TC error,1 in which only the combined error can be assessed. Within this combined error will be some level of compensation between the two individual errors, due to the tuning to maximize TC climatology performance. The tuning within individual GCMs has led to multiple TC definitions that vary between detectors, models and model parameterizations (e.g., Vitart et al. 1997; Yokoi and Takayabu 2009; Yokoi et al. 2009), model resolution (e.g., Bengtsson et al. 1995; Murakami and Sugi 2010), and TC basins within individual models (e.g., Camargo and Zebiak 2002).

With so many TC detection variants, it can be difficult to compare results between studies. To address this issue, Walsh et al. (2007) derived a relationship for wind speed thresholds as a function of grid resolution. Their method reduces the overall subjectivity of the wind speed threshold choice and has been reported to work well in some studies (e.g., the multiple grid resolution (20–180 km) Japanese Meteorological Agency/Meteorological Research Institute (JMA/MRI) studies of Murakami and Sugi 2010). However, for the application of the threshold to the Commonwealth Scientific and Industrial Research Organisation (CSIRO) Conformal Cubic Atmospheric Model (CCAM) system, a selection of phase 3 of the Coupled Model Intercomparison Project (CMIP3) global climate models, and European Centre for Medium-Range Weather Forecasts (ECMWF) Interim Re-Analysis (ERA-Interim) data by D. Abbs (2011, personal communication), a subjectively determined reduction of 0.7 was found to be necessary to give acceptable results. This highlights the difficulty in developing a truly objective wind speed threshold. However, the Walsh et al. (2007) approach of using a semiobjective resolution-dependent wind speed threshold does provide a benchmark for comparison between models and studies, which is likely to continue to be important for some time. Even the highest-resolution climate models (e.g., Oouchi et al. 2006; Chauvin et al. 2006; Zhao et al. 2009; Murakami and Wang 2010; Murakami et al. 2011, and references therein) cannot sufficiently resolve the TC eyewall wind structure for realistic wind speed thresholds to be applied.

The TC detection scheme used in Bengtsson et al. (2007) avoided grid dependency by degrading data to a common coarse resolution before applying a relative vorticity threshold. A warm-core requirement identified TCs and tropical depressions, which was later refined (Strachan et al. 2013) to reduce the number of non-TC, monsoon-like features being identified. It is not clear from their results how well the refined warm-core definition distinguishes between tropical depressions and TCs.

b. OWZP TC detector

Thresholds are applied to the Okubo–Weiss–Zeta (OWZ) diagnostic to identify regions of enhanced vorticity with low (or weak) deformation. These are combined with moisture and wind shear thresholds (T13a) to give the OWZ predictor (OWZP). This novel TC detector uses only large-scale variables that can be resolved in the coarser-resolution climate models, which enables the use of uniform thresholds. To address the issue of error compensation between the model and detector error the detector is developed and tuned to ERA-Interim data at a resolution similar to contemporary climate models and individually verified against observed TCs. If a circulation is present in reanalysis data that matches an observed TC, then it will likely have a coarse grid signature similar to that in a coarse grid climate model.

We are not aware of other published studies that develop and tune the TC detector in reanalysis data prior to its application to climate model data. While Bengtsson et al. (2007) and Strachan et al. (2013) applied their detectors to reanalysis data, they were developed independently of the reanalysis data. The detector of Murakami and Sugi (2010) was developed mostly independent of reanalysis data, with only the 850-hPa relative vorticity threshold tuned to reproduce the observed global annual TC numbers. Murakami and Sugi (2010) retuned their detector for each model grid resolution they presented. While the TC detection scheme of Klepek et al. (2008) was tuned to 40-yr ECMWF Re-Analysis (ERA-40) data, it was only tuned and applied to the North Atlantic basin and was not applied to climate models. The latter two papers include simple performance measures, which we apply in this paper to provide some form of performance comparison.

---

1. Here the GCM modeled error refers to the GCM TC climatology error (i.e., too many or too few and incorrect spatial and temporal distributions).
Limited comparison is also made with Strachan et al. (2013).

The OWZ diagnostic (described below) essentially replaces the absolute vorticity used in traditional detectors and genesis indices. Its development (T13b) was inspired by the low-deformation vorticity observed at the center of wave-relative closed circulations in which observed and modeled TCs have been found to form (Dunkerton et al. 2009; Wang et al. 2009, 2010; Montgomery et al. 2010). It is in this region that the vortex upscale cascade and system-scale intensification, which underlies the bottom-up TC formation theory (Hendricks et al. 2004; Montgomery et al. 2006; Tory et al. 2006, 2007), is believed to be most efficient (Dunkerton et al. 2009; T13b). Furthermore, T13b use simple axisymmetric theory to argue that system-scale vortex spinup is most efficient in regions of highly rotating flow in solid-body rotation, which is identifiable by enhanced values of OWZ. OWZ is based on the Okubo–Weiss parameter, normalized by the square of the vertical component of relative vorticity $\zeta$,

$$\text{OW}_\text{norm} = \left[ \frac{k^2 - (E^2 + F^2)}{\chi^2} \right].$$

Here, $E = (\partial u / \partial x) - (\partial v / \partial y)$ and $F = (\partial u / \partial x) + (\partial v / \partial y)$ are the square of the stretching and shearing deformation, respectively. Only positive values of $\text{OW}_\text{norm}$ are considered in the OWZ, which represents a range of vortical flows from linear shear or flows with deformation exceeding vorticity ($\text{OW}_\text{norm} = 0$) to solid body rotation ($\text{OW}_\text{norm} = 1$). OWZ is a measure of the vertical component of absolute vorticity ($\eta = \zeta + f$) weighted by $\text{OW}_\text{norm}$ and multiplied by the sign of Coriolis $f$ to ensure positive values for cyclonic vorticity in both hemispheres,

$$\text{OWZ} = \eta \times \max(\text{OW}_\text{norm}, 0) \times \text{sign}(f).$$

Thus, for solid body rotation (zero flow deformation) OWZ has the same magnitude as the absolute vorticity, and the ratio of OWZ to $\eta$ decreases as the flow deformation increases.

T13b hypothesized that all TCs form in regions of enhanced low- to midtroposphere OWZ. They found 95% of observed TCs in the 20-yr period from 1989 to 2008 could be associated with regions of enhanced OWZ in ERA-Interim data on both the 850- and 500-hPa pressure levels, and for about 90% of TCs this enhanced OWZ was present for at least 24 h prior to the observed TC being declared to have reached a sustained wind speed of 17 m s$^{-1}$. T13b concluded that these numbers provide good support for the hypothesis, especially when taking into account imperfect observations and reanalysis data.

In addition to the OWZ thresholds, minimum relative humidity thresholds on the 950- and 700-hPa pressure levels, minimum specific humidity thresholds on the 950-hPa level, and maximum wind shear thresholds between the 850- and 200-hPa levels are applied. About 85% of observed TCs were matched to ERA-Interim circulations that satisfied these thresholds for a minimum of 48 h. With these arbitrarily chosen thresholds, the system essentially has a miss rate of 15% and a false alarm rate of about 50%. In the current paper, a tuned version of the OWZP is presented as a TC detector, and the TC detector performance applied to the same 20 yr of ERA-Interim data are presented. The identical OWZP TC detector is applied to a sample of climate models from CMIP3 in a companion paper (Tory et al. 2013a, hereafter T13a).

In section 2, the various datasets and methods used in this study are presented, and results are described in section 3. Finally, a discussion and summary of the work are given in section 4.

2. Data and methodology

a. TC data and definitions

TC data from the International Best Track Archive for Climate Stewardship (IBTrACS) database (e.g., Knapp et al. 2010) is used for verification purposes. The database contains a comprehensive compilation of quality-controlled global TC best-track data sourced from various meteorological organizations and agencies around the world (available online at http://www.ncdc.noaa.gov/oa/ibtracs). Included in the IBTrACS database are TC location, maximum sustained wind, and central pressure at 6-hourly intervals (0000, 0600, 1200, and 1800 UTC) throughout the lifetime of the TC and for many storms before and after the system has first reached TC intensity. For the purpose of this study, we only consider data at 0000 UTC in order to be consistent with the temporal resolution of the climate model data examined in subsequent papers. Here we also objectively define a TC as any system in the IBTrACS database that reached the 10-min maximum sustained wind speed of at least 17 m s$^{-1}$ during any 0000 UTC period of its lifetime; the first such position where the wind speed exceeds 17 m s$^{-1}$ is considered to be the genesis location. Only data that we could directly verify in ERA-Interim are used in this paper, which limited the study to the 20-yr period of 1989–2008.

b. ERA-Interim

The atmospheric data were obtained from the ECMWF reanalysis products (ERA-Interim) for the period 1989–2008. ERA-Interim data used here have a horizontal...
resolution of $1.5^\circ \times 1.5^\circ$ available globally and four times daily. The wind fields required for the detection process are the zonal and meridional wind components used to compute (i) OWZ on the 850- and 500-hPa pressure levels, (ii) vertical shear of the horizontal wind (abbreviated to vertical wind shear) between the 850- and 200-hPa levels, and (iii) a TC steering velocity at the 700-hPa level. Other fields required are relative humidity at the 950- and 700-hPa levels and specific humidity at the 950-hPa level. To minimize unavoidable grid “hard wiring” problems it was decided that the detection technique should be developed and applied at the same grid resolution. Thus, both the ERA-Interim data and the climate model data in T13a are interpolated to a $1^\circ \times 1^\circ$ grid.

The reanalysis data can be thought of as a best grid-based estimate of the state of the atmosphere. Because neither the reanalysis nor IBTrACS data are perfect, it would be expected that some observed TCs might not be identified in the reanalysis data. Serrano (1997) applied two vortex detection schemes to ERA-40 and found the two schemes identified circulations that matched about 75%–85% of observed TCs and that one scheme detected about 5%–7% more than the other. This latter result is a reminder that no TC detection scheme is perfect. While many of the missed storms may be due to the failure of the reanalysis system to reproduce a circulation surrounding a tropical cyclone, some fraction will be due to deficiencies in the detection algorithms. The OWZP method described here was designed and tuned within the ERA-Interim to avoid as much as possible preconceived ideas of what a coarsely resolved TC should look like. This is one of the main reasons for developing a detection scheme that searches for the large-scale environment that TCs form in, rather than searching for TCs directly.

c. Detection of tropical cyclones

A detailed description of the TC detection and tracking algorithm can be found in T13b. Here we give an overview of the method in context of detecting TC genesis frequency in the ERA-Interim data. The detection and tracking algorithm is a multistep process. Circulations of interest are first identified using a set of “initial” thresholds that mostly tests for grid points with enhanced OWZ on both the 850- and 500-hPa pressure levels. Next, neighboring grid points are grouped together into clumps that represent individual storms at a particular time. A storm tracking routine assesses which clumps belong to the same storm and concatenates them into individual storm tracks. At each position along the storm track, the storm is assigned individual OWZ, relative humidity (RH), vertical wind shear, and specific humidity (SH) values, which are subjected to additional sets of “core” thresholds and conditions to decide whether the storm should be considered a TC. Over 100 combinations of core thresholds were tested, and results from the best performing combination are presented in this paper (Table 1, criterion 1), as well as two other combinations to illustrate the threshold sensitivity.

1) INITIAL THRESHOLDS

The initial thresholds are specified to identify primary circulations with dynamic potential to support TC formation. The conclusion in T13b that most if not all TCs form where the OWZ is enhanced throughout much of the lower to middle troposphere underlies our choice to use OWZ thresholds on the 850- and 500-hPa pressure levels as the basis of the initial detection. After a subjective analysis of two seasons of the western North Pacific (WNP) and South Pacific (SP) basins and one season of the North Atlantic (NA) basin, we chose the initial OWZ thresholds of 50 and 40 $10^{-8}$ s$^{-2}$ for the 850- and 500-hPa levels, respectively. To simplify the computation, additional weak vertical wind shear and relative humidity thresholds were added to eliminate any circulations that were clearly not tropical cyclones. As can be seen in Table 1, these thresholds are deliberately set to conservative values to ensure no tropical systems are eliminated during this initial stage.

2) CORE THRESHOLDS AND CONDITIONS

Each position along the storm track (associated with a clump of grid points that satisfied the initial thresholds) is labeled “true” if its individual clump thresholds satisfy the more rigorous core thresholds and conditions and “false” if not. The rigorous conditions are based on the results of T13b, which showed that the rigorous thresholds should generally be satisfied for 48 h or more before the circulation should be considered a TC. Three consecutive 0000 UTC true values thus rate as a valid TC detection. This simple condition can be problematic in coastal regions and for storm tracks that may temporarily fail to satisfy the rigorous thresholds. After a number of subjective tests we arrived at the following conclusions: (i) if less than two grid points in the storm clump are over water the storm is considered to be land affected and the clump reassigned a false value and (ii) there must be three

### Table 1. Threshold criteria used for TC detection in ERA-Interim data. OWZ is given in units of $10^{-8}$ s$^{-2}$, RH in percentage, 850–200-hPa vertical wind shear (VWS) in m s$^{-1}$, and SH in g kg$^{-1}$.

<table>
<thead>
<tr>
<th>Criteria</th>
<th>OWZ$_{500}$</th>
<th>OWZ$_{500}$</th>
<th>RH$_{950}$</th>
<th>RH$_{700}$</th>
<th>VWS</th>
<th>SH$_{950}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Initial</td>
<td>50</td>
<td>40</td>
<td>70</td>
<td>50</td>
<td>25</td>
<td>—</td>
</tr>
<tr>
<td>Criterion 1</td>
<td>60</td>
<td>50</td>
<td>85</td>
<td>70</td>
<td>12.5</td>
<td>12.3</td>
</tr>
<tr>
<td>Criterion 2</td>
<td>50</td>
<td>40</td>
<td>85</td>
<td>70</td>
<td>12.5</td>
<td>12.3</td>
</tr>
<tr>
<td>Criterion 3</td>
<td>60</td>
<td>40</td>
<td>85</td>
<td>70</td>
<td>12.5</td>
<td>12.3</td>
</tr>
</tbody>
</table>
TABLE 2. Mean tropical cyclone frequency per year globally, in two hemispheres and in the six basins defined in Fig. 3. The respective mean values for the entire globe, two hemispheres, and the six basins are statistically similar at the 95% significance level (as determined using the bootstrap sampling; null hypothesis being that the annual mean values of the observed and ERA-Interim TCs for respective regions are statistically different). Also shown is the normalized mean bias (%) between the two datasets.

<table>
<thead>
<tr>
<th>Basin</th>
<th>Global</th>
<th>SH</th>
<th>NH</th>
<th>SP</th>
<th>SI</th>
<th>NI</th>
<th>WNP</th>
<th>ENP</th>
<th>NA</th>
</tr>
</thead>
<tbody>
<tr>
<td>Observations</td>
<td>78.7</td>
<td>24.4</td>
<td>54.4</td>
<td>9.1</td>
<td>15.3</td>
<td>3.7</td>
<td>24.3</td>
<td>14.2</td>
<td>12.2</td>
</tr>
<tr>
<td>ERA-Interim</td>
<td>81.5</td>
<td>26.4</td>
<td>55.1</td>
<td>9.5</td>
<td>17.0</td>
<td>5.0</td>
<td>27.8</td>
<td>12.5</td>
<td>9.9</td>
</tr>
<tr>
<td>Bias (%)</td>
<td>4</td>
<td>8</td>
<td>1</td>
<td>4</td>
<td>11</td>
<td>35</td>
<td>14</td>
<td>−12</td>
<td>−19</td>
</tr>
</tbody>
</table>

consecutive time periods in which the thresholds are satisfied (i.e., three consecutive true values), regardless of the storm history. It is possible that a more sophisticated set of conditions could be devised to improve the TC detection performance of storms impacted by land, but the time-consuming nature of further experimentation brought diminishing returns.

The TC genesis location is taken to be the point at which the criteria are satisfied for the third consecutive 0000 UTC time. Given that a reasonable proportion of TCs are observed to reach our TC intensity definition after two or four consecutive 24-h time periods, the OWZP specification of genesis after three consecutive time periods will introduce some temporal and spatial error along the track for those storms.

3. OWZP TC detector performance

In this section TCs detected in the ERA-Interim data are compared with the IBTrACS observed TCs to evaluate the performance of the OWZP method. The performance of the detection scheme in reproducing the observed TC formation climatology is first investigated. This is followed by a storm by storm performance assessment using a number of categorical statistical skill measures. In the latter assessment, both missed detections (misses) and false detections (false alarms) are included in the error assessment, whereas they effectively negate one another in the former. This level of performance assessment is unprecedented in the TC detection literature. In addition few studies have applied TC detectors to reanalysis data, which means comparison opportunities are limited.

a. Detector performance comparisons

The only TC detectors we can compare the OWZP performance to are those of Murakami and Sugi (2010), Kleppke et al. (2008), and to a lesser extent Strachan et al. (2013). The Murakami and Sugi detector was tuned in one of their experiments to best reproduce the annual number of TCs globally in the Japanese 25-yr Reanalysis (JRA-25) reanalysis data from 1979 to 2003. They binned the genesis locations into $5^\circ \times 5^\circ$ boxes and calculated the Taylor skill score (Taylor 2001), which evaluates the TC spatial distribution. They returned an encouraging value of 0.96 (a perfect score is 1.0). For comparison, we repeated the calculation for the OWZP TC detector applied to ERA-Interim data from 1989 to 2008 and returned a Taylor skill score of 0.95, which demonstrates a very similar level of performance. However, the results we present in the remainder of this section suggest the Taylor skill score applied to the 20-yr climatology alone is not a rigorous test of the detection system. The Kleppke et al. detector, which was tuned to ERA-40 and applied only to the North Atlantic basin, appears to have overpredicted annual TC numbers by more than 60% (see their Fig. 4) during the period 1998–2001. For the same period in the North Atlantic basin, the OWZP detection in ERA-Interim underpredicts TC numbers by about 19% (section 3c).

Strachan et al. (2013) applied their detector to three reanalysis datasets with broad ranging results (see their Fig. 1). It is perhaps unfair to compare the performance of their detector with the OWZP detector because the former has not been tuned to optimize performance with any of the reanalysis datasets. However, the Strachan et al. detector applied to ERA-Interim data at first glance appears to show very good performance, with a very small bias (estimated from their Fig. 1) for global TC numbers. The global bias of the OWZP detector (Table 2) is of very similar magnitude. On the other hand, the Strachan et al. detector biases in hemispheric TC numbers are much larger: about $-20\%$ and $+30\%$ for the Northern and Southern Hemispheres, respectively, compared with $+1\%$ and $+8\%$ for the OWZP detector (Table 2).

b. Geographical distribution of TC genesis positions

A direct comparison of detected and observed TC genesis locations in Fig. 1 suggests that the OWZP TC detection method reproduces the observed global TC genesis climatology very well, a result consistent with the Taylor skill score value of 0.95. The distributions of TC genesis positions, expressed in terms of a mean annual frequency for each 2.5° latitude and longitude band (Fig. 2), also demonstrate good agreement between the observed and the ERA-Interim TCs. The
The spatial distribution of TC genesis positions (Fig. 2a) is well reproduced over the entire globe, except for a slight poleward bias (~2°) of the genesis maxima in ERA-Interim detected TCs over the Northern Hemisphere. Similarly, the meridional distribution of TC distribution is generally well represented for the two hemispheres (Figs. 2b,c).

The kernel density estimation (KDE; Bowman and Azzalini 1997) is applied to offer a more objective assessment of the geographical distribution of the observed and ERA-Interim TCs within the six TC basins defined in Fig. 3. The KDE is a nonparametric technique whereby a density function or kernel can be used to construct a smooth probability density estimate of the spatial data (see, e.g., Ramsay et al. 2008; Chand and Walsh 2009). For convenience, the KDEs here are displayed (Fig. 4) with contours that enclose proportions of genesis positions corresponding to the median and to quartiles such that the outermost contour contains 75% of the total genesis positions and the middle and innermost contours contain 50% and 25%, respectively. A comparison of the observed and detected KDE contours provides a clear indication of the differences in geographical distribution.

On the whole, the spatial distribution of the observed TC genesis positions compare very favorably with those detected in the ERA-Interim data for all TC basins. In the SP basin (Fig. 4a), for example, the maximum density region enclosed by the 25% KDE contour over the Coral Sea region compares well between the two datasets, with only slight variations in structure of the other two contours. Similarly, TCs forming over the main development region in the NA basin (Fig. 4f) are well represented, although the densities in the Gulf of Mexico and immediately east of Florida are lower, indicating an underprediction bias in these regions. In the south Indian (SI) basin the distribution is good, although a slight overprediction bias is evident between about 90° and 120°E. Comparisons of genesis patterns in other basins show good performance except over the Bay of Bengal (~20°N, 75°–90°E) in the north Indian (NI) basin (Fig. 4c), where a poleward shift in the TC genesis location is noted. This shift is partly responsible for the poleward shift in the genesis maxima in the ERA-Interim detected TCs for the Northern Hemisphere, as noted earlier in Fig. 2a. It is difficult to ascertain here whether the discrepancy is due to the detection method or if it is related to different observing practices in this basin. A similar shift is evident in Bengtsson et al. (2007, their Fig. 5) and Strachan et al. (2013, their Fig. 4f), which suggests different observing practices may be responsible.

c. Mean annual number of TCs

The mean annual number of detected and observed TCs are compared over the period 1989–2008, globally, hemispherically, and in the individual basins. Overall, the mean annual number of TCs in the two datasets is
everywhere statistically similar at the 95% significance level (Table 2). The significance test is described in the appendix. There is a slight overestimation of the mean number of TCs in the SP and the NI basins, with the latter perhaps partially due to different observing practices in the region (noted above). The OWZP detector also slightly overestimates TCs in the WNP. This overestimation, although statistically insignificant, could be due to the inclusion of some of the long-lived tropical depressions that frequent the WNP (e.g., Chen et al. 2006). The underestimation in the NA and eastern North Pacific (ENP) basins may be related to the higher frequency of smaller TCs that form in these basins that may be less well resolved in the reanalysis data. Three TCs were detected in the South Atlantic basin compared to one TC in the IBTrACS database that reached our 17 m s\(^{-1}\) TC definition. No TCs were detected or observed in the south east Pacific region (east of 120°W) over the 20-yr period (e.g., Fig. 1), in agreement with observations.

![Figure 2](image1.png)

**FIG. 2.** Distribution of TC genesis positions expressed in terms of a mean annual frequency for each 2.5° (a) latitude band, (b) Southern Hemisphere longitude band, and (c) Northern Hemisphere longitude band. Solid lines represent observed TCs, and broken lines represent TCs detected in ERA-Interim data.

![Figure 3](image2.png)

**FIG. 3.** The six TC basins used in the study: SP (0°–40°S, 130°E–120°W), SI (0°–40°S, 30°–130°E), NI (0°–40°N, 40°–100°E), WNP (0°–40°N, 100°E–180°), ENP (0°–40°N, 180°–100°W), and NA (0°–40°N, −100°W–0°).
FIG. 4. Genesis locations of (left) observed and (right) OWZP detected TCs in the six basins defined in Fig. 3, together with the KDE contours 25, 50, and 75 containing 25%, 50%, and 75% of the total genesis locations, respectively; the innermost contours represent 25%.
d. Interannual variability

The mean annual distribution of observed and detected TCs (Fig. 5) is used to investigate basinwide interannual variability. These time series demonstrate that the OWZP detection method captures the annual TC frequency quite well, and shows mostly good performance in reproducing the interannual variability. An objective measure of this performance is presented in Table 3, which shows the correlation coefficients and the normalized root-mean-square error (NRMSE) between the two datasets for all basins. The NRMSE and correlations are quite encouraging for all basins but the NI. No statistically significant correlation exists between the observed and detected datasets in the NI basin, whereas correlations are significantly high in the other basins. The correlation coefficients range from 0.53 for the SI ocean to 0.94 for the NA. The NRMSE is about 3 times greater in the NI basin than all other basins.

The weaker performance in the NI basin is consistent with earlier suggestions that either there is something unique about TC formation in that basin or differences in observing practices. Murakami et al. (2012) retuned their TC detection scheme over the NI basin to deal with this problem.

e. Case-by-case analysis

The above results demonstrate very good performance of the OWZP method in reproducing the climatological distribution, annual frequency, and interannual variability of TCs in ERA-Interim data. However, these climatological comparisons do not rigorously test the OWZP method as a TC detector, because they do not assess correct detection or nondetection of individual storms.

Table 3. Correlation coefficients R and the NRMSE between the observations and ERA-Interim for all TC basins. NRMSE is RMSE normalized by the mean number of observed TCs. Statistically significant correlations, determined using the bootstrap method at the 95% significance level, are highlighted in bold. Also given is the CV for the two datasets.

<table>
<thead>
<tr>
<th>Basin</th>
<th>SP</th>
<th>SI</th>
<th>NI</th>
<th>WNP</th>
<th>ENP</th>
<th>NA</th>
</tr>
</thead>
<tbody>
<tr>
<td>NRMSE</td>
<td>0.24</td>
<td>0.22</td>
<td>0.74</td>
<td>0.21</td>
<td>0.27</td>
<td>0.22</td>
</tr>
<tr>
<td>R</td>
<td>0.85</td>
<td>0.53</td>
<td>0.02</td>
<td>0.63</td>
<td>0.55</td>
<td>0.94</td>
</tr>
<tr>
<td>CV (Obs)</td>
<td>0.43</td>
<td>0.17</td>
<td>0.48</td>
<td>0.20</td>
<td>0.30</td>
<td>0.38</td>
</tr>
<tr>
<td>CV (ERA)</td>
<td>0.42</td>
<td>0.20</td>
<td>0.34</td>
<td>0.15</td>
<td>0.23</td>
<td>0.44</td>
</tr>
</tbody>
</table>

2 The root-mean-square error here has been normalized by the associated mean number of observed TCs to account for the differences in annual TC numbers between basins.
To address this issue, we examine the number of hits, misses, and false alarms using a contingency table (Table 4) and calculate associated statistical skill measures to measure the OWZP performance overall and within individual basins. Note that without an observed database of nondeveloping TCs it is not possible to verify the correct rejection contingency of Table 4, and it cannot be incorporated in the statistical assessment.

A number of categorical statistical measures (e.g., McBride and Ebert 2000) are computed from the elements of the detected–observed TC contingency table. These include the probability of detection (POD; or hit rate),
\[ \text{POD} = \frac{H}{H + M}. \]
The POD is the number of hits divided by the total number of TCs observed; thus, it provides a simple measure of the number of TCs successfully detected.

The false alarm ratio (FAR) is the number of false alarms divided by the total number of TCs detected and it provides a simple measure of the ratio of false detections to the total number of detections,
\[ \text{FAR} = \frac{F}{H + F}. \]
A bias score (BIAS) is also computed in which the total number of TCs detected is divided by the total number of TCs observed,
\[ \text{BIAS} = \frac{H + F}{H + M}. \]

The critical success index (CSI; or threat score) is another statistical measure that takes into account errors associated with both hits and misses. It is appropriate for data not dominated by misses and false alarms (e.g., Schaefer 1990). The CSI ranges from 0 to 1, with a value of 1 indicating a perfect score, and is given by
\[ \text{CSI} = \frac{H}{H + F + M}. \]

Table 5 shows the above statistical measures for the detected and observed TCs over the period 1989–2008. Of the total number of TCs observed globally, 78% are correctly detected and 25% are falsely detected. The global number of false alarms and misses are similar, yielding a bias ratio of ~1, which is not surprising given the thresholds were tuned to maximize performance in all basins. Similar results are found for the two hemispheres and the SP. The other basins have biases between about 10% and 20%, except for the NI basin, where the FAR is particularly high. The CSI for the NI basin is also relatively low compared to the other basins, indicating that the OWZP TC detector has difficulty correctly identifying NI basin TC circulations listed in the IBTrACS database.

To get an indication of how maximum observed intensity affects the detections, we recalculated the POD for a variety of observed maximum wind speeds (Table 6). For the borderline TCs (maximum sustained wind speed between 17 and 19 m s\(^{-1}\)), the detection rate was as low as 48%, whereas 93% of hurricane strength TCs (maximum sustained wind speeds of >32 m s\(^{-1}\)) were detected. The low detection rate for the weaker storms is perhaps

---

**Table 4.** Schematic of the 2 × 2 contingency table for verification of categorical data. The correct rejection cannot be calculated in this study because of the absence of a database of observed nondeveloping systems.

<table>
<thead>
<tr>
<th>Detected</th>
<th>Yes</th>
<th>No</th>
</tr>
</thead>
<tbody>
<tr>
<td>Yes</td>
<td>Hit (H)</td>
<td>False alarm (F)</td>
</tr>
<tr>
<td>No</td>
<td>Miss (M)</td>
<td>Correct rejection</td>
</tr>
</tbody>
</table>

**Table 5.** Categorical statistical skill measures of ERA-Interim TC detections over the period 1989–2008. Total number of storms correctly detected (hit), falsely detected (false alarm) and missed (miss), plus categorical statistical skill measures of ERA-Interim TC detections over the period 1989–2008.

<table>
<thead>
<tr>
<th>Basin</th>
<th>Hit</th>
<th>False alarm</th>
<th>Miss</th>
<th>POD</th>
<th>FAR</th>
<th>BIAS</th>
<th>CSI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Global</td>
<td>1229</td>
<td>401</td>
<td>345</td>
<td>0.78</td>
<td>0.25</td>
<td>1.04</td>
<td>0.62</td>
</tr>
<tr>
<td>SH</td>
<td>387</td>
<td>141</td>
<td>100</td>
<td>0.79</td>
<td>0.27</td>
<td>1.08</td>
<td>0.62</td>
</tr>
<tr>
<td>NH</td>
<td>842</td>
<td>260</td>
<td>245</td>
<td>0.77</td>
<td>0.24</td>
<td>1.01</td>
<td>0.63</td>
</tr>
<tr>
<td>SP</td>
<td>135</td>
<td>54</td>
<td>47</td>
<td>0.74</td>
<td>0.29</td>
<td>1.04</td>
<td>0.57</td>
</tr>
<tr>
<td>SI</td>
<td>252</td>
<td>87</td>
<td>53</td>
<td>0.83</td>
<td>0.26</td>
<td>1.11</td>
<td>0.64</td>
</tr>
<tr>
<td>NI</td>
<td>49</td>
<td>51</td>
<td>25</td>
<td>0.66</td>
<td>0.51</td>
<td>1.35</td>
<td>0.39</td>
</tr>
<tr>
<td>WNP</td>
<td>435</td>
<td>120</td>
<td>51</td>
<td>0.90</td>
<td>0.22</td>
<td>1.14</td>
<td>0.72</td>
</tr>
<tr>
<td>ENP</td>
<td>198</td>
<td>51</td>
<td>86</td>
<td>0.70</td>
<td>0.20</td>
<td>0.88</td>
<td>0.59</td>
</tr>
<tr>
<td>NA</td>
<td>160</td>
<td>38</td>
<td>83</td>
<td>0.66</td>
<td>0.19</td>
<td>0.81</td>
<td>0.57</td>
</tr>
</tbody>
</table>

---

3 Because of the requirement that all core thresholds and conditions must be satisfied for three consecutive 0000 UTC time periods, shorter-lived TCs are more likely to be missed than a longer-lived TCs. This introduces an underprediction bias in formation regions close to land.
not surprising, because (i) the more difficult to detect short-lived storms are more likely to be of low intensity and (ii) there is likely to be more subjectivity in the wind speed data entries around the critical tropical cyclone/tropical storm threshold.

f. Sensitivity tests

In developing the OWZP most of the threshold fine tuning was performed on the new and untested OWZ. Less tuning was required for the relative humidity and vertical wind shear components of the OWZP, because appropriate threshold ranges could be determined from other studies [e.g., Nolan (2007) and Paterson et al. (2005), respectively]. The 12.5 m s\(^{-1}\) wind shear threshold is also consistent with the recent study of McGauley and Nolan (2011), who found about 85% of TCs formed in wind shear less than 11 m s\(^{-1}\). The specific humidity threshold was able to be tuned independently of the other thresholds to remove the handful of higher-latitude false detections because it has minimal impact on tropical circulations.

To illustrate the sensitivity of the OWZ thresholds to the OWZP two additional threshold combinations are introduced labeled criteria 2 and 3 in Table 1. Criteria 2 and 3 include a reduction of 10 \(\times 10^{-6}\) s\(^{-1}\) in the 500-hPa OWZ threshold relative to the tuned thresholds (criterion 1, Table 4). The 850-hPa OWZ threshold is reduced only in criterion 3 by 10 \(\times 10^{-6}\) s\(^{-1}\), which matches the initial OWZ thresholds (section 2c) and those used in the preliminary OWZ development described in section 4 of T13b.

Figure 6 shows the relative differences between the NRMSE and CSI for the three OWZ combinations are quite small, with the exception of the NI basin NRMSE. While these case by case statistics suggest the OWZP performance is relatively insensitive to the choice of OWZ threshold, the sensitivity to the basinwide mean annual TC numbers is perhaps a little more pronounced (Fig. 7). Both Figs. 6 and 7 show that criterion 1 demonstrates the better performance on the whole, whereas criterion 2 slightly outperforms criterion 1 in the SP, ENP, and NA basins. The poorer performance of criterion 3 is due to overprediction in all basins.

Overall, the numbers of TCs detected using criterion 1 are statistically similar to the corresponding observations (at the 95% significance level) in all basins (Fig. 7).

These statistically similar results, however, are not present in the SI, NI, and WNP basins for criteria 2 and 3.

4. Discussion and summary

a. OWZP TC detection method

The OWZP TC detector was constructed using only large-scale atmospheric variables that can be resolved in coarse-resolution numerical models. This approach was taken to minimize the need for subjective TC definitions (i.e., wind speed thresholds) that tend to vary with grid resolution (e.g., Walsh et al. 2007) and vary between models and in some studies between TC basins within the same model (e.g., Camargo and Zebiak 2002). The development and tuning in reanalysis data enable a more rigorous testing of the method itself, compared with TC detectors that have been developed and tuned in climate model data, because the former can be verified against actual observed TCs, whereas the latter can only be verified against TC climatology. Furthermore, there is no way to separate errors associated with a model’s ability to correctly reproduce the TC climatology, and errors associated with the TC detection technique. Thus, any tuning to maximize performance in

---

**Table 6.** Probability of detection rate for TCs of varying intensity where the intensity is the maximum observed 0000 UTC 10-min mean wind speed.

<table>
<thead>
<tr>
<th>Category (m s(^{-1}))</th>
<th>17–19</th>
<th>19–21</th>
<th>21–32</th>
<th>17–32</th>
<th>&gt;32</th>
</tr>
</thead>
<tbody>
<tr>
<td>POD</td>
<td>0.48</td>
<td>0.64</td>
<td>0.57</td>
<td>0.69</td>
<td>0.64</td>
</tr>
</tbody>
</table>
reproducing TC climatology will result in some degree of compensation between the two errors. For example, a model that has trouble reproducing large-scale processes critical to TC formation (e.g., African easterly waves in the North Atlantic) typically underpredicts TC formation. Tuning to maximize performance in such models will inevitably reduce the TC detection thresholds to improve the TC climatology by detecting additional circulations that would otherwise not be considered TCs.

The OWZP development and tuning in reanalysis data, at similar resolution to present-day climate models has ensured only atmospheric variables resolvable in climate model data are included in the parameter. Furthermore, the cumulus parameterization used at these grid resolutions, which is critical to the construction of the model TC-like circulations (e.g., Tory et al. 2006), should generate very similar TC-like vortex structures in both the reanalysis and climate models. A known exception is moisture differences between various cumulus parameterization schemes used in reanalysis and climate models [e.g., the dry bias of the National Centers for Environmental Prediction (NCEP) reanalyses; Bony et al. 1997]. Applying moisture thresholds tuned in a model using one cumulus parameterization scheme to a model with different cumulus scheme could produce a systematic bias in TC detections.

Despite including only atmospheric variables resolvable in coarse-resolution data, it is unlikely that all grid dependency will be avoided. To minimize this problem we first chose to interpolate all data to a common grid ($1^\circ \times 1^\circ$) to avoid problems associated with grid-dependent hardwiring (e.g., the two grid point neighbor limit, section 2b). Then we chose OWZ thresholds that define an environment conducive to TC formation, rather than thresholds that define the TC-like circulation, to avoid formation thresholds that vary with TC size due to resolution limitations.

b. OWZP TC detection performance

The main purpose of this paper is to document the performance of a novel TC detector. Our motivation has been to develop a technique that avoids some of the known limitations of traditional TC detection techniques. The OWZP method will of course be subject to other limitations, but what exactly they are and how problematic they are will not be apparent until the method is applied to climate models (see T13a). All TC detection techniques will be subject to performance limitations, but
the extent of these limitations is largely unknown because of minimal published performance testing results of other TC detectors. In this paper, the OWZP performance is documented in sufficient detail for it to be used as a benchmark for future performance comparison and for comparison with other techniques. In the absence of similar performance results for traditional TC detection methods we are not in a position to recommend one technique over another. The similar level of performance between the OWZP and Murakami and Sugi TC detectors from the minimal comparison we were able to make suggests that the OWZP TC detection could at the very least be used to complement existing techniques. The application of a suite of TC detection schemes to climate models will help reduce projection uncertainties, in the same way that analyzing a suite of models reduces uncertainty.

While there is variation between basins in all statistical measures (Tables 2, 3, and 5 and Fig. 6), the NI basin stands out as performing particularly poorly in nearly all measures. Either there is something unique about NI basin TCs or perhaps a systematic difference in observing practice exists compared to other basins. Of the circulations identified as TCs there is a higher formation density near the Indian and Bangladesh coastline than observed (Fig. 4c), a result also evident in Fig. 5 of Bengtsson et al. (2007). For storms heading toward India and Bangladesh, this difference could be due to faster development of TCs in the region or conservative TC declarations by forecasting staff to maximize warning time for the particularly vulnerable people of the region. Because of the uniqueness of the NI results, we exclude the NI basin from the remainder of the discussion.

It is clear from the normalized bias (Table 3) that the OWZP TC detector overpredicts TCs in some basins (SI and WNP) and underpredicts in others (ENP and NA), with only the SP basin registering a minimal bias. The overpredicting basins tend to have more larger and longer-lived TCs with long tracks over open water (not shown) that are more easily detected by the OWZP, whereas the underpredicting basins have a higher number of smaller storms and perhaps more storms influenced by land (not shown). The smaller ERA-Interim OWZP formation density of the Caribbean and Gulf of Mexico compared with that observed (Fig. 4f) could possibly be due to more faster forming but shorter-lived TCs in the region. The biases could also be partially due to regional differences in observing techniques (e.g., Landsea 1993; Emanuel 2000; Buckley et al. 2003).

With respect to interannual variability the two highly correlating basins (NA and SP) happen to be basins with relatively high coefficients of variation (CV; Table 3), in which case random error in the OWZP detected storms will have less impact on the year-to-year trends than the basins with smaller correlations and smaller CV (SI, WNP, and ENP). The normalized RMSE varies little across most basins, ranging from 0.21 to 0.24 for the WNP, SI, NA, and SP, with only the ENP a little larger at 0.27. These consistent numbers should be expected for a system tuned to minimize error across all basins.

With respect to correct detections of individual storms, the overpredicted basins have higher false alarm rates (FAR) than miss rates (one minus the POD) and vice versa for the underpredicted basins. The previously discussed basin statistics (normalized bias, normalized RMSE, and correlation) do not account for all incorrect storm identification, because the error is based on the difference between the false alarms and misses per basin per year (i.e., error 1 of Fig. 8). The critical success index (Table 5), on the other hand, sees all incorrect storm identification as errors (i.e., error 2 of Fig. 8). The global CSI of 0.62 is perhaps less flattering than the OWZP technique deserves, because there is no reward for correct detection of nondeveloping systems (i.e., correct rejection, Table 4). Interestingly, while the best performing basins with respect to the error 1 statistics were shared between basins (smallest bias SP; smallest NRMSE WNP, SI, and NA; and highest correlation NA) only the...
WNP performs significantly better than the global average (16% higher CSI) for the error 2 statistic. The SI basin CSI is similar to the global average, and the SP, ENP, and NA basins have a CSI of about 5%–8% lower than the global average.

The standout better CSI performance of the WNP basin raises the question of whether the overall performance of the OWZP could have been improved with tuning to optimize the CSI in all basins. While the relaxing of thresholds to increase the number of detections might be expected to increase the performance of the underpredicting basins (NA and ENP), the sensitivity tests show only the relaxed thresholds of criterion 2 lead to a small performance improvement in only the NA basin, despite the criterion 2 NRMSE being slightly smaller in the SP, ENP, and NA basins (Fig. 6). These small differences between criteria 1 and 2 CSI and NRMSE demonstrate the low sensitivity of the OWZP performance to the choice of OWZ thresholds, which is reassuring because it signifies a degree of stability in the method that will be important when applied to a wide variety of climate models.

In closing, we note that the OWZP method is capable of realistically detecting TC genesis frequency in the ERA-Interim data. In T13a, the scheme is applied directly to a selection of CMIP3 climate models without modification to the detection thresholds or any form of downscaling. The present and future climate TC climatologies are assessed and preliminary projected changes in TC frequency based on the few selected models are documented. The combined work gives us confidence that the OWZP TC detector, with the objectively determined, grid-resolution-independent thresholds should match traditional detector performance, and should become a valuable TC detection technique for direct detection of TCs in coarse-resolution climate model data.

Acknowledgments. We thank Andrew Dowdy and Xingbao Wang for their valuable insight. We acknowledge the Pacific Climate Change and Science Program (PCCSP) project for supporting this work. PCCSP is funded by AusAID, in collaboration with the Department of Climate Change and Energy Efficiency, and delivered by the Bureau of Meteorology and the Commonwealth Scientific and Industrial Research Organisation (CSIRO).

APPENDIX

Statistical Significance Tests

Statistical significance tests, for evaluating the difference between the mean values of TCs in the observations and the ERA-Interim data, are conducted using the bootstrap resampling method (Efron and Tibshirani 1991). This method is based on the assumption that the empirical density function associated with the 20 yr of data is a reasonable estimate of the unknown population density function. Concerning this study, we resample the two sets of data separately 1000 times and calculate the associated mean statistics for each bootstrap sample, giving a total of 1000 bootstrap distribution of sample means for each set of data. The two distributions are then compared using their respective 95% confidence intervals. If bootstrap confidence intervals for the two distributions overlapped, then the mean of the two sets of data are considered statistically similar.

REFERENCES
