How well are Tropical Cyclones represented in reanalysis data sets?

How Well are Tropical Cyclones Represented in Reanalysis Data Sets?

Kevin Hodges*

Department of Meteorology, University of Reading, Reading, United Kingdom

Alison Cobb

Department of Meteorology, University of Reading, Reading, United Kingdom

Pier Luigi Vidale

National Centre for Atmospheric Science, Department of Meteorology, University of Reading, Reading, United Kingdom

**Corresponding author address:* Department of Meteorology, University of Reading, Reading, United Kingdom.

E-mail: k.i.hodges@reading.ac.uk

ABSTRACT

Tropical cyclones (TCs) are identified and tracked in six recent reanalysis data sets and compared with those from the IBTrACS best track archive. Results indicate that nearly every cyclone present in IBTrACS over the period 1979-2012 can be found in all six reanalyses using a tracking and matching approach. However, TC intensities are significantly under-represented in the reanalyses compared to the observations. Applying a typical objective TC identification scheme, it is found that the largest uncertainties in TC identification occur for the weaker storms; this is exacerbated by uncertainties in the observations for weak storms and lack of consistency in operational procedures. For example, it is unclear whether certain types of storms, such as tropical depressions, subtropical cyclones and monsoon depressions, are included in the best track data for all reporting agencies. There are definite improvements in how well TCs are represented in more recent, higher resolution reanalyses; in particular MERRA2 is comparable with the NCEP-CFSR and JRA55 reanalyses, which perform significantly better than the older MERRA reanalysis.

1. Introduction

Tropical cyclones (TCs) are one of the most damaging weather-related natural hazards on the planet, causing 42 % of the United States catastrophe-insured losses in the period 1992-2011 (King 2013). Individual intense events can result in severe losses. For example, Hurricane Katrina resulted in an estimated death toll of 1,833 people and financial losses of over $125 billion (Adeola and Picou 2014). Weaker storms, such as tropical depressions can also have an impact in terms of loss of life and disruption in vulnerable societies (for the Caribbean 2009). It is therefore important to utilise the available data and new analysis techniques to better understand their properties and behaviour, with the aim of mitigating their societal, economic and environmental impacts.

Due to the relatively short observational record of TCs, and problems with sampling within the record, there is considerable uncertainty in the variability of TCs in terms of frequency, over climate time scales of the last 100 years (Landsea 2007; Landsea et al. 2009), resulting in uncertainty in the interannual variability and trend detection. The use of reanalyses to detect TCs provides an opportunity to reduce this uncertainty (Truchelut et al. 2013), by allowing the creation of a larger data sample which, when used in conjunction with the historic observational data, can help to provide more confidence in TC numbers than the observations alone. Reanalyses combine observations with a short forecast from a general circulation model (GCM) to produce gridded data sets with regular output intervals constrained by the observations, and can act as a bridge between the observations of TCs and simulated tempestology. However, there can be problems in using reanalyses related to the changing observing system, in particular the introduction of spurious trends (Bengtsson et al. 2004a), and the fact that different reanalyses use different GCMs with different parameterizations and different data assimilation methods, all of which can contribute to differences between them. The study of Schenkel and Hart (2012) previously considered the representation of TCs in the northern hemisphere in several reanalyses, including several of those used in this study, by manually tracking the best track TCs in the reanalyses, and found considerable variation in the properties of TCs between the reanalyses, for location, and a consistently large underestimate of intensity (10m winds and Mean Sea Level Pressure) for all the reanalyses. This uncertainty in the representation of TC properties in reanalyses can introduce uncertainty into their detection in these data, so that detection criteria are often tailored to the particular reanalysis of interest (Murakami 2014).

Another motivation for a careful study of the properties of TCs as represented by reanalyses is that they are often used as a means of calibrating TC detection and tracking schemes before applying them to climate models (Bengtsson et al. 2007a). This is done by first applying the detection to the reanalyses or operational analyses and adjusting the detection criteria to give similar numbers of TCs to those found in the observations, provided by the TC warning centers best track data. This may be problematic if there are large differences between how reanalyses represent TCs in terms of their properties, such as structure and intensities, or if there are biases in the best track data. The model dynamical core, parameterizations and resolution all play a critical role in determining the output of extreme events in reanalysis data. These vary widely, with in particular newer generations of reanalyses being produced at higher resolutions and with more modern data assimilation systems. For climate models, the IPCC 5th Assessment (2013) stated that there is medium evidence and high agreement that year-to-year count variability of Atlantic hurricanes can be well simulated by modest resolution (100 km or finer) atmospheric GCMs (AGCMs) forced by observed Sea Surface Temperatures (SSTs). Both Strachan et al. (2013) and Roberts et al. (2015) show that 60 km is adequate for simulating interannual variability, although not intensity. Recent work by Murakami (2014) showed that, when considering five reanalyses (also included in this study), the highest resolution reanalysis was not always the best in terms of simulating the TC climatology and properties, nor did the higher-resolution reanalyses produce significantly more intense storms than those with lower resolutions, suggesting that the simulation of TCs in the reanalyses is highly dependent on model formulation (Schenkel and Hart 2012) and/or data assimilation strategy. However, if we can understand the uncertainties of TCs in the reanalyses, they may provide a useful means of extending the observations, for example, by extending their lifecycles to include the extratropical transition (Jones et al. 2003) and beyond, which would be useful for TC related extratropical risk analysis and GCM assessment (Haarsma et al. 2013). The use of reanalysis could also assist in the identification of subtropical and hybrid tropical storms (Roth 2002; Guishard et al. 2009), which are also associated with severe weather, providing a more complete set of tropical storm data for use in GCM assessment than is perhaps currently present in best track data; the inclusion of these types of storms in the best track data sets is highly variable between the operational centres.

The main aim of this paper is to quantify the uncertainties in how well TCs are represented in a number of recent reanalyses, and how this affects the objective identification of TCs in reanalyses. This is achieved by exploring:-

1. how well reanalyses represent the observed TCs in the best track data using direct track matching.

2. how well does an objective identification scheme identify the best track TCs in the reanalyses and what might be the cause of differences.

2. Data and Methods

Data from six recent reanalyses are used in this study and described below. Also used are best track data produced by the tropical warning centers as post season analyses of the TC tracks. These have been combined into the International Best Track Archive for Climate Stewardship (IBTrACS) data set (Knapp et al. (2010)) and are used in this study for verifying the TCs identified in the reanalyses. The IBTrACS-ALL, which includes data from all agencies, is used in this study. The common period of 1979-2012 is used throughout for all data sets, except for one reanalysis where the period is 1980-2012. Throughout the rest of the paper the following nomenclature is used; the term Tropical Cyclone (TC) is used for warm core storms generally and, where appropriate, the term Tropical Storm (TS) is used for TCs with wind speeds greater than 17 m s^{-1}.

a. Best Track dataset

For full details of the IBTrACS-ALL data set, see Knapp et al (2012). The original wind speed data in knots is converted to wind speed in m s^{-1}. The World Meteorological Organization (WMO) standard for reported tropical cyclone wind speed is maximum 10-minute sustained winds at 10 m height over a smooth surface; however, this is rarely observed, therefore some discrepancy between agencies is apparent. Different agencies apply different wind-averaging periods, with the East Pacific, North Atlantic (RSMC Miami), and central Pacific (RSMC Honolulu) using 1-minute averaging periods, North Indian (RSMC New Delhi) using a 3-minute period and the other agencies using 10-minute averaging periods (Schreck III et al. 2014). The 10-minute wind speeds are converted to 1-minute wind speeds using a factor of 1.13, which has traditionally been used (Harper et al. 2010), and the data from RSMC Miami and New Delhi are used in their original form. However, there are uncertainties in the accuracy and fidelity of this conversion, with different conversion factors for at-sea, off-sea, off-land and in-land parts of the storm suggested (Harper et al. 2010). Other uncertainties also exist in the best track data, which have been discussed is several studies; a summary of these uncertainties can be found in the appendix of Hodges and

Emerton (2015). They include issues relating to location and intensity uncertainties and operational differences between agencies. This is further discussed in the Discussion section.

For the analysis of the identified TCs in different ocean basins the IBTrACS basin boundaries (Knapp et al. 2010) have been used, with TCs assigned to a particular ocean basin, based on where the storm reaches maximum wind speed intensity.

b. Reanalysis datasets

Meteorological centers around the world produce reanalysis data sets as an ongoing enterprise. The reanalyses are essentially based on frozen operational numerical weather prediction (NWP) systems. New reanalyses are often released following significant improvements in the models and data assimilation schemes. The reanalyses differ in terms of the models and data assimilation methods used to produce them, therefore differences in their output are to be expected. Six recent global atmospheric reanalysis data sets have been analysed for TCs in this study and are summarized in Table 1. They include the European Centre for Medium-Range Weather Forecasts (ECMWF) Interim reanalysis (ERAI) (Dee et al. 2011); the Japanese 25-year reanalysis (JRA25) (Onogi et al. 2007) and 55-year reanalysis (JRA55) (Kobayashi et al. 2015); the National Aeronautics and Space Administration (NASA) Modern-era Retrospective Analysis for Research and Applications (MERRA)(Rienecker and coauthors 2011) and the following version 2 (MERRA2) (Bosilovich and coauthors 2015; Molod et al. 2015); and the National Centers for Environmental Prediction (NCEP) Climate Forecast System Reanalysis (NCEP-CFSR) (Suranjana and Coauthors 2010). The NCEP-CFSR reanalysis is the only coupled atmosphere-ocean-land surface-sea ice reanalysis. The NCEP-CFSR, MERRA and MERRA2 all use different versions of the 3D variational (3D-Var) data assimilation scheme: the Grid-point Statistical Interpolation (GSI) scheme (Shao et al. 2016). For MERRA and MERRA2 the Incremental Analysis Update (IAU) (Bloom et al. 1996; Rienecker and coauthors 2011) system is also used. The data period used for all the reanalyses is 1979-2012, except for MERRA2, which starts in 1980. A key difference between the JMA reanalyses and the reanalyses produced by the other agencies is the assimilation of tropical wind retrievals (TWR). Wind profile data over and around tropical cyclone centers are retrieved from historical data and processed and assimilated as if they were dropsonde observations (Hatsushika et al. 2006). With the integration of this additional wind data, the intensity of the storms in the JMA reanalyses is found to be improved (Hatsushika et al. 2006). Another difference between the reanalyses is that the NCEP-CFSR uses a technique to improve the representation of TCs by adjusting the location of the tropical vortex to its observed location before the assimilation of storm circulation observations (Suranjana and Coauthors 2010). The MERRA2 reanalysis also uses this method. All the reanalyses in this study make use of quality control processes and bias correction for the diverse range of observations that are assimilated, for example, the variational bias correction of satellite radiances (Dee and Uppala 2009).

c. Tropical cyclone detection method

The analysis of TCs in this study relies on identifying and tracking them. The first step is to track all tropical disturbances, in both hemispheres, before applying two different identification methods to separate the TCs from other tropical systems. This is different from some other schemes where the identification is performed during the tracking and hence only identifies the TC stage of the lifecycle. Though not crucial to this study, the approach taken here identifies much more of the lifecycle, including the precursor and post extratropical transition stages (Jones et al. 2003).

For the first step, where all systems in the domain are tracked, the tracking methodology is based on Hodges (1994, 1995, 1999). The domain extends to 60N in the NH and 60S in the SH. The tracking method uses the 6 hourly relative vorticity at the levels 850, 700, 600hPa, vertically averaged. This data is spectrally filtered using triangular truncation to retain total wavenumbers 6-63. The spectral coefficients are also tapered to further smooth the data using the filter described in Sardeshmukh and Hoskins (1984). The spectral filtering acts to remove the noise associated with the smallest spatial scales in the vorticity, which produces more reliable tracking in data of this type, and to remove the large scale background, which is also found to be beneficial. The tracking proceeds by identifying the off-grid vorticity maxima, by applying a maximisation scheme (Hodges 1995), that exceed a value of $5 \times 10^{-6} s^{-1}$ in each time frame (SH scaled by -1). These are initially linked together using a nearest neighbor approach and then refined by minimizing a cost function for track smoothness, subject to adaptive constraints on displacement distance and track smoothness (Hodges 1999). The use of the vertically averaged vorticity is different from some previous studies using this tracking algorithm, where the single level of 850hPa vorticity reduced to T42 resolution was used (Strachan et al. 2013; Roberts et al. 2015; Bell et al. 2013; Bengtsson et al. 2007b; Manganello et al. 2012). The use of the vertically averaged vorticity is found to improve the temporal coherency when a vorticity maximum shifts between levels (Serra et al. 2010; Fine et al. 2016) and results in more of the system lifecycle being detected. A simple vertical average is found to be sufficient, even though the levels are not evenly spaced, since, once spectrally filtered, there is little difference from using the mass weighted vertical average. Only tracks that last at least 2 days (8 time steps) are retained for further analysis. Whilst observed TCs can have lifetimes shorter than 2 days, this only covers the period when they are determined to be TCs, whereas the tracking scheme used here aims to identify the precursor and post-TC stages resulting in much longer lifetimes (see Figure 1c and d) so that using the 2 day threshold is not detrimental to detecting nearly all the observed TCs in the reanalyses, as shown below in the results section.

Previous methods used to detect TCs in reanalysis or GCM data rely on applying particular criteria, representative of the properties of TCs, such as some thresholds on intensity, e.g. Mean Sea Level Pressure (MSLP) minima, low level wind intensities or vorticity extrema, and a threshold on the warm core structure either determined directly as a temperature anomaly, or inferred from the presence of decreasing winds or vorticity between the lower and upper troposphere, for example Bengtsson et al. (1995) and related methods. These are often applied as part of the tracking scheme itself, which is different from the approach used here. A minimum period of one day is typically imposed, for which these criteria are satisfied contiguously, and that they are satisfied only over the ocean by imposing the land-sea mask. The criteria based on intensity and structure can be strongly dependent on the model resolution and how processes important to TC development, such as convection, microphysics and surface drag, are represented in the model. This has resulted in some studies using resolution dependent identification criteria (Walsh et al. 2007; Manganello et al. 2012) or tuning the identification criteria to maximize the detected TCs, for example in reanalyses compared with observations (Murakami 2014), and some studies have used basin dependent criteria (Camargo et al. 2005). The study of Horn et al. (2014) has shown that the subjective choice of different identification criteria is the main reason for differences between the numbers of TCs identified by different identification schemes.

In this study a dual approach is taken to isolate the TCs from all the tracked systems. Taking the tracks identified in the first stage, where all systems are tracked, the first approach used to isolate the TCs is used to see which of the observed TCs in the IBTrACS data set can be found in the reanalyses, without applying any criteria dependent on intensity or structure. This approach makes use of spatio-temporal matching: a track in the reanalyses matches with a track in IBTrACS if the mean separation distance between them, computed over the time period that they overlap, is less than 4^0 (geodesic), and is the least mean separation distance if more than one track satisfies this criterion, where any amount of temporal overlap is allowed. This will be termed the "direct matching" method. A similar approach has previously been used for extra-tropical cyclones (Hodges et al. 2003). The relaxed criterion on the temporal overlap is chosen because, in general, the TCs in IBTrACS have much shorter lifetimes compared to the tracks in the reanalyses produced by the tracking scheme. Several diagnostics are produced from the matched tracks, such as the mean separation distance distribution, lifetime distribution and intensity distribution based on low level winds, 10m and 925hPa, and MSLP.

The second approach used to isolate the TCs from all the tracked systems is to objectively identify them using a typical set of identification criteria based on intensity and structure; this will be termed the "objective detection" method. The criteria used are similar to those used previously with this tracking algorithm (Bengtsson et al. 2007a,b; Strachan et al. 2013). This requires adding additional fields to the tracks, namely the T63 vorticity at levels 850, 700-200hPa to provide intensity and warm core criteria. This is done by recursively searching for a vorticity maximum at the different levels using the maximum at the previous level as a starting point for a steepest ascent maximization applied to the B-spline interpolated field. A search radius of 5^0 (geodesic) is used centered on the location at the previous level. The same approach is used in the Southern Hemisphere (SH) by multiplying fields by -1. Also added are the Mean Sea Level Pressure (MSLP) minimum and maximum winds at 10m and 925hPa as alternative measures of TC intensity. For MSLP a steepest descent method is used with the B-spline interpolation and a search radius of 5^0 (geodesic) centered on the tracked vorticity center to find the closest pressure minimum, whilst for the winds a direct search for the maximum winds within 6^0 of the tracked center is used. The criteria for identification are:

1. the T63 relative vorticity at 850 hPa must attain a threshold of at least $6 \times 10^{-5} s^{-1}$.

2. the difference in vorticity between 850 and 200 hPa (at T63 resolution) must be greater than $6 \times 10^{-5} s^{-1}$ to provide evidence of a warm core.

3. the T63 vorticity center must exist at each level between 850 and 200hPa for a coherent vertical structure.

4. criteria (1) to (3) must be jointly attained for a minimum of 4 consecutive time steps (one day) and only apply over the oceans.

5. tracks must start within 30S to 30N.

The approach used here means that the tracking and identification is performed at a common resolution for all the reanalyses, making the tracking and identification as resolution independent as possible, although the actual model resolution will still have some impact on the identification.

The TCs identified by the objective detection method are also matched against the observed tracks in IBTrACS, using the same criteria as in the direct matching method, to determine the hit and miss rates of the identification scheme.

The tracking is applied to each full year, January-December, for the Northern Hemisphere (NH) and July to June the following year in the Southern Hemisphere (SH), resulting in 34 years in the NH and 33 in the SH (33 and 32 respectively for MERRA2).

3. Results

In this section the ability of the different reanalyses to simulate different aspects of TC behavior is assessed and compared to the observed TC activity, as represented by the IBTrACS database described in the Best Track dataset subsection.

a. Direct Matching Results

The number of TCs in IBTrACS that match with a storm in the reanalyses for each reanalysis using the direct matching method are summarized in Table 2 for both NH and SH. This shows that $\sim$ 95% of the TCs in IBTrACS are identified in the reanalyses in the NH and $\sim$ 92% in the SH. The different reanalyses are remarkably similar in this respect. In general the TCs not found in the analyses tend to be the weakest and/or short lived TCs in IBTrACS in both hemispheres. Some of the missing TCs fail to pass the 2 day lifetime threshold imposed on the reanalysis tracks. There is also some evidence that the number of missing TCs in the reanalyses, according to the matching criteria, are reduced in the later period after 2000: compared to the earlier period, the number of matches increases to $\sim$ 98% in both NH and SH. This improvement may be associated with the assimilation of improved observations, in particular the availability of surface scatterometer winds from the QuikSCAT satellite data from mid-1999 until the end of 2009 and continuing with similar data from other remote sensing platforms since then.

To see how the TCs identified in the reanalyses by the direct matching method compare with those in IBTrACS several sets of statistics are produced.

1) LOCATION

Figure 1a and b show distributions for the mean separation distance (geodesic distance) between the identical reanalysis tracks and those of IBTrACS, obtained using the direct matching method, in the NH and SH respectively. In the NH (Figure 1a) the majority of TCs identified in the reanalyses have a mean separation from those in IBTrACS of less than 2^0 (220km), with the peak of the distribution for each reanalysis typically at less than 1^0 (110km). The smallest mean separation distances occur for JRA55, with the distribution peak at 0.5^0 (56km) and the largest for MERRA, with the distribution peak at 1^0 and the other reanalysis somewhere in between. The

JRA55 separation distances are comparable with those from the much higher resolution (T1279; 16km)) operational analyses of ECMWF (Hodges and Emerton 2015) (Appendix), which may be a consequence of the assimilation of the TWR observations in JRA55. This conjecture is strengthened by the fact that JRA25, which also assimilates TWR data, is comparable in terms of the mean separation distances to the much higher resolution NCEP-CFSR reanalysis. It is also apparent that MERRA2 has improved over MERRA with respect to the separation distances. In general the mean separation results for the NH (Figure 1a) are consistent with those found by Schenkel and Hart (2012) for the identical reanalyses considered. In the SH (Figure 1b) a rather similar picture is seen, with each of the reanalyses occurring in the same order as in the NH of best to worse. Whilst the separation distances appear slightly larger for some reanalyses in the SH, i.e. ERAI and MERRA, the others are comparable with the results in the NH, highlighting the improvement in the SH in the more recent reanalyses compared with older reanalyses.

2) LIFETIME

Figure 1c and d show the lifetime distributions in the NH and SH respectively. In the NH it is apparent that the TCs identified in the reanalyses have much longer lifetimes than the TCs in the observations. This is a consequence of not imposing any criteria during the tracking to identify TCs. Imposing the TC detection criteria during the tracking would truncate the tracks to the TC stage alone, and would introduce a dependency of the liftime on the chosen criteria and how well TCs are represented in the reanalyses in terms of intensity and structure. The extended lifecycles include pre-TC stages, e.g. easterly waves and the stage after extratropical transition. Some of the reanalysis TCs can exist for longer than one month, in which time a precursor disturbance can travel across an ocean basin, develop into a TC and recurve to high latitudes undergoing extratropical transition, whereas none of the observed TC tracks last this long. The distributions for the different reanalyses are quite close together, showing that rather similar lifetimes are obtained for all the reanalyses. A similar set of results is obtained in the SH, although the distributions for the reanalyses are a little noisier, due to the smaller number of observed TCs in this hemisphere.

3) Latitude of Maximum Intensity

The latitude at which the maximum intensity is attained in terms of the 10m winds is shown for the NH and SH in Figure 1e and f respectively. In the NH the distributions show that, whilst most TCs in the reanalyses attain their maximum intensity at similar latitudes to those in the observations, there are some TCs that attain their maximum intensity at much higher latitudes. A possible cause for this behavior is that, because of the longer lifecycles that are identified in the reanalyses, some storms only attain their maximum intensity as they recurve to higher latitudes and become larger and better represented at synoptic scales. Whilst this could be addressed by restricting the reanalysis tracks to just the TC stage, this would mean either truncating the tracks where they overlap with the best track data (Hodges and Emerton 2015), or using the detection criteria based on intensity and structure discussed above to define the TC part of the lifecycle. Either of these approaches introduces a degree of subjectivity: the first as it depends on the different operational practices of the operational agencies, and the second because it depends on how well TCs are represented in the different reanalyses. Also, for this part of the study, we want to see what exactly is in the reanalyses in terms of TC lifecycle and restricting the lifecycles defeats this objective. This is also important for future work, such as studies of extratropical transition and risk associated with TCs and their later lifecycle stages in extratropical regions. A similar situation may also occur for the TC stage itself, where the relatively low resolution of the reanalyses means that TCs are not well represented at the small spatial scales of TCs in the tropics, but become better represented as they move to higher latitudes. A similar picture is seen for the SH (Figure 1f). This type of behaviour is often seen for TCs identified in relatively low resolution climate model simulations (Manganello et al. 2012).

4) Intensity

Also examined are the maximum intensity distributions of the TCs for three intensity measures: minimum MSLP and maximum 10m and 92hPa wind speeds, which are shown in Figure 2 for both NH and SH TCs. For both MSLP (Figure 2a, b) and 10m wind speeds (Figure 2c, d) in the NH and SH it is clear that all the reanalyses underestimate the intensity of TCs compared to the observations and that the intensities are model dependent. This is not surprising considering the relatively low spatial resolutions of the reanalyses where the assimilation of observations cannot correct for this. Previous studies with dynamical downscaling of individual historical TCs, such as Katrina, have shown that resolutions $\sim$1-5km with a non-hydrostatic model are necessary to simulate TC inner-core processes correctly in order to enable the right magnitude of wind intensities (Davis et al. 2008) to be simulated. However, some studies using hydrostatic models with parameterized convection at resolutions $\sim$10km can certainly produce TCs with depths as large if not larger than observed (Manganello et al. 2012). Coupling to the ocean has also been found to be important in correctly simulating TC intensity (Kilic and Raible 2013), although only the NCEP-CFSR reanalysis applies any such coupling and its impact on the reanalysis and TCs is uncertain.

The results for intensity based on the MSLP (Figure 2a, b) show that in general the more recent reanalyses, NCEP-CFSR, JRA55 and MERRA2 have deeper TCs; this is more evident in the SH, although in both hemispheres few TCs reach depths below 940hPa. The more recent reanalyses may be performing slightly better with respect to this intensity measure, possibly due to better use of the available observations and improved models, and not necessarily due to resolution. For

10m wind speeds (Figure 2c, d), much larger differences are seen between the different reanalyses, although, as already mentioned, none of them can simulate the strongest intensities seen in the observations. NCEP-CFSR has the most intense TCs in terms of 10m wind speeds, with some TC almost attaining intensities of 50 ms^{-1} (Category 3 TS) but with no Category 4 or 5 (Saffir-Simpson scale) TSs. The weakest maximum 10m wind speed intensities are produced by the MERRA reanalysis with no TCs surpassing 30 ms^{-1}, which barely reaches Category 1 TS. However, the more recent MERRA2 reanalysis shows a significant improvement being comparable with the JRA55 reanalysis in having TCs that can almost attain 10m wind speeds of 40 ms^{-1} (Category 1 TS), although less than that seen for the NCEP-CFSR reanalysis. The results for the reanalyses TC 10m wind speeds show similar behavior in both hemispheres. The results for both 10m wind and MSLP maximum intensities are generally consistent with those of Schenkel and Hart (2012) for the NH. One problem with using the 10m winds from the reanalyses is that they are not a direct model prognostic field, but are computed as a diagnostic, though not necessarily in the same way for each reanalysis. They are generally computed as an extrapolation from the lowest model level to the surface using profile functions and corrected when over land for terrain roughness to conform to the World Meteorological Organization (WMO) standard for SYNOP observations (for example see, ECMWF (2015)). However, for some reanalyses this is not done for the actual analyses: for example in MERRA, it is performed during the IAU cycle, so does not see the full analysis increment, and is an average over four model time steps (private communication, Michael Bosilovich, NASA). To evaluate the uncertainty further, the maximum wind speeds at the 925hPa pressure level associated with the TCs are also considered (pressure level winds are obtained by interpolation between model levels); the TC 925hPa winds are shown in Figure 2e, f for the NH and SH respectively. The downside to using the 925hPa winds is that there are no available observations with which to compare with, although this is not critical here, where we just want to see if the same differences between the reanalyses, as seen for 10m winds, occur at this level. The results for the wind speed intensity at 925hPa show a rather different perspective from those at 10m, with both NCEP-CFSR and MERRA2 having comparable values in the tail of the distribution with values as high as 60 ms^{-1}. The MERRA reanalysis is now comparable with the other reanalyses of JRA55, JRA25 and ERAI.

5) Wind Speed-Pressure Relationship

The wind speed-pressure relationship is often used by the operational centers to estimate winds from pressure measurements and surface pressure from wind measurements, for which various quadratic empirical relationships have been developed based on cyclostrophic balance (Knaff and Zehr 2007). Hence, the wind-pressure relationship of TCs is often considered in studies of TCs in models and reanalyses (Roberts et al. 2015) to compare with the observed relationship, although it should be noted that the observations may themselves be estimated from one of the empirical relationships, which can differ between agencies (Knaff and Zehr 2007). Figure 3a shows the wind-pressure relationship for the observations and the TCs identified in the different reanalyses using the direct matching method in the NH. The wind-pressure relationship is determined using the 10m wind speeds and MSLP values, by determining the maximum attained 10m wind speed and taking the MSLP value at the same time. The results show that all the reanalysis reflect the underestimate of both the 10m wind speeds and MSLP depths of the TCs, this being most prominent for MERRA. This can be related to the radius of maximum wind (RMW), computed for the reanalyses at the time of maximum 10m wind intensity, and shown for the NH in Figure 3c. The RMW is not available for all the agencies in IBTrACS but we estimate it at the time of maximum wind intensity, based on the simple Rankine model described by Knaff and Zehr (2007), this gives RMW values for the observations predominately below $\sim$100km (1^0) and a peak around $\sim$50km

(0.5^0). This is consistent with the findings of Kimball and Mulekar (2004) for North Atlantic TSs who made use of an extended "best track" data set. For all the reanalyses the RMW are seen to be too large (Figure 3c). Assuming gradient wind balance for the TCs, and the fact that RMWs are too large and wind intensities are too low for the reanalyses implies that the pressure difference between the storm centers and the environment is also too low, consistent with the wind speed-pressure relationship in Figure 3a. The fact that the NCEP-CFSR has the strongest wind intensities and one of the smallest RMWs is also consistent with the result in Figure 3a that NCEP-CFSR is closest to the observed wind speed-pressure relationship, whereas MERRA, which has the weakest maximum wind speeds and large RMWs, is the worst of the reanalyses in this respect. MERRA2 shows a significant improvement over MERRA in terms of the wind speed-pressure relationship which can be understood in terms of the improved maximum wind speeds and lower RMWs. In fact, MERRA2 has the lowest RMWs, although is not as strong in intensity (10m wind speed) as NCEP-CFSR.

The fact that NCEP-CFSR appears to perform the best in terms of the wind speed-pressure relationship may be the result of the vortex relocation scheme used by the NCEP-CFSR assimilation system, which, as pointed out by Schenkel and Hart (2012), will result in improved vortex location, which in turn may lead to improved TC intensities as a result of the TC being in the correct environment. Allied to this, Schenkel and Hart (2012) also pointed out that observations within the TC vicinity are less likely to be rejected by the assimilation scheme, due to smaller differences with the first-guess field. However, the situation is likely more complex than this, as MERRA2 also uses the vortex relocation method and has the lowest RMWs but is not the most intense in terms of wind speed. JRA55, on the other hand, with a similar resolution to MERRA2, has the smallest location errors (Figure 1a, b), does not use vortex relocation, but does assimilate best track data as synthetic dropsondes (Hatsushika et al. 2006) and has comparable intensities to MERRA2 and a wind speed-pressure relationship, also very similar to MERRA2. Hence, it appears that there are complex trade-offs occurring within the assimilation systems. In the SH the wind-speed pressure relationship (Figure 3b) and RMWs (Figure 3d) appear to be very similar to those in the NH: in particular the wind-speed pressure relationship appears to be closely associated with the ordering of the 10m wind speeds of the reanalyses shown in Figure 2b.

b. Objective Identification

Following the assessment of how well TCs are represented in the chosen reanalyses it is of interest to see how existing objective TC identification schemes perform in order to try and understand the impacts of the differences between reanalyses on objective TC identification. This is important, as objective schemes are the only way to identify TCs in climate model simulations and they are often contrasted with reanalyses as a means of verification at comparable resolutions. As Murakami (2014) has shown, detection schemes have to be tuned to particular reanalyses to optimally detect TC/TS frequencies. This is also what tends to happens in operational settings, where detection schemes are often tuned to a particular operational setup, so that applying them to data from a different operational center can give very different numbers of detected TCs from the in-house method (c.f. Fig. 22 of Kobayashi et al. (2015)). Some schemes also adjust identification criteria by ocean basin (Camargo and Zebiak 2002) to account for model biases. However, these are not appealing approaches in the climate model context, where a fixed set of criteria, applied in a common resolution framework, will provide a better comparison between different model simulations or different climate scenarios (Shaevitz and Coauthors 2014). To assess how one such scheme performs, the objective detection method described in the methodology section, based on the vorticity at multiple levels between 850 and 200hPa, is applied to the vorticity tracks obtained from the tracking of all vorticity centres.

1) Annual Counts

The annual average TC counts are determined for each ocean basin (Figure 4) and are shown in Figure 5. In the NH the annual number is in reasonably good agreement with the observations of IBTrACS apart from MERRA, which has ∼30 fewer identified TCs, whilst the other reanalyses are slightly over or under in number, a result also previously noted by Murakami (2014) using the same criteria. However, in the SH the identification has resulted in a much higher number than in the observations, which occurs for all the ocean basins. The overestimation is particularly large in the South Pacific (SP) region; the South Atlantic (SA) region also has more identified systems than are in the observations. These differences will be discussed further in the Discussion section.

2) Matching Against IBTrACS

To further analyse the objectively identified TCs, they are matched against the observed TCs of IBTrACS using the direct matching method to identify the common storms between the two and the false positive and negative detections. The results of this track matching are shown in Table 2 in terms of the Probability of Detection (POD) and False Alarm Rate (FAR). The POD is defined here as the number of matched storms for each reanlysis divided by the total number of storms in the observations, and the FAR by the number of non-matched storms in each reanalyses divided by the total number of storms in the same reanalysis. Also shown in Table 2, for comparison, are the POD for the direct matching results, before applying the objective criteria, discussed in the Direct Matching Results subsection, which shows an almost uniform detection rate 0f 0.95 across all the reanalyses in both hemispheres, although this is lower in the SH than the NH. The reason why the POD for the SH is lower for the pre-criteria matching is likely related to differences in the observations that are assimilated in the reanalyses between the two hemispheres, as there is no dependence on structure or intensity for detection for these results.

For the POD based on using the objective detection method the values are much lower, with the best detection for JRA55 and the worst for MERRA in both hemispheres, although POD is higher in the SH than the NH, possibly due to differences in sample sizes. The FAR (Table 2) shows values ranging from 0.16 for JRA25 to 0.36 for NCEP-CFSR in the NH. The fact that JRA25 has the lowest FAR may be related to this reanalysis having the lowest resolution, hence, detecting fewer small scale and possibly weaker storms; this could be investigated using GCMs of varying resolution. In the SH, FAR is much higher, as might be expected from the previous discussion, due mostly to the higher number of TCs detected compared with the observations. From these values of POD and FAR it is apparent that, although similar numbers of TCs are detected in the NH using the objective detection method, they need not be identical to the ones in the observations.

To explore the POD and FAR values in more detail the storms that are in the observations and that match and do not match with those identified in the re-analyses, using both identification methods, pre-objective direct matching and post-objective matching, are further analyzed relative to their attained category in the observations according to the Saffir-Simpson scale determined from the 1min. observed winds. Hence, the IBTrACS storms are partitioned into the categories according to the 1min. winds before matching them against the reanalysis tracks, as previously described. Since different agencies use different wind intensity scales, this approach provides a more consistent classification across the different ocean basins. Since some weak storms in IBTrACS have no wind information, they are excluded from this analysis; Murakami (2014) excluded tropical depressions from their study, although it is unclear how this is achieved for the reanalyses, apart from applying the agency wind thresholds. The results of this analysis by category are shown in Tables 3 and 4 for the NH and SH respectively. In the NH, Table 3 shows that for the objectively identified TCs it is the weakest categories that have the poorest level of matches between the reanalyses and IBTrACS, in particular for the tropical depressions, although many tropical depressions in

IBTrACS are excluded due to lack of wind information. However, for the TS category (between tropical depression and Category 1) the best performing reanalyses at this level, JRA25 and JRA55, match with 78.5% of IBTrACS storms, while for the worst performing (MERRA) only 41.6% of IBTrACS storms match. For the higher TS wind speed categories the percentage of matches with IBTrACS steadily increases with category on progressively smaller sample sizes, i.e. 92%, 98%, 99.5% and 100% for CAT1-CAT5 respectively, for the best performing JRA25 and JRA55 and considerably worse for MERRA (63.5, 75, 83, 82.5, 92%) with NCEP-CFSR and MERRA2 comparable with JRA25 and JRA55. Re-calculating the POD for just Cat1-Cat5 TS (Table 5) the best performing reanalyses, JRA25 and JRA55 now have values 0.95.

In the SH, Table 4 shows that a fairly similar situation occurs as in the NH for the objectively identified TCs, except that it is apparent there are virtually no tropical depressions available to compare with in the observations, either because very few of this category of storms have any wind values, or more likely that they are not generally included in the best track data sets in this hemisphere; this is discussed further in the discussion section. The best degree of matches again occurs for the JRA25 and JRA55, ranging from 84-89% for the weakest TSs (TS Category) to 95% for CAT5.

The POD, for CAT1-CAT5 objectively identified TS only, shown in Table 5, shows that for this intensity range the values are comparable in both hemispheres and comparable with the results in the study of Murakami (2014) who restricted their study to this intensity range, although they used different skill metrics compared to here and in the study here there is no special tuning of the objective detection parameters for each reanalysis as in Murakami (2014).

For the TCs identified using the direct matching method (pre-objective), previously discussed in the Direct Matching Results subsection, the matching by observation category (not shown) indicates consistently high POD values as reported in the Direct Matching Results subsection for all categories and reanalyses.

To understand the nature of the TCs, identified by the objective detection method, in the reanalyses that do not match with the IBTrACS TCs, in particular in the SH, those that do not match are binned according to the latitude of their genesis. For the SH this is shown in Figure 6(a). This shows essentially two groups of storms: those with genesis within 0-20S and those with genesis occurring south of 20S. The genesis for all TCs in IBTrACS is almost entirely within 0-20S (not shown). Examining these two groups on non-matching objectively identified TCs separately, a scan of the tropical storm advisories (discussed later) indicates that some of the identified storms in the first group can be found in the advisories but not IBTrACS; this is discussed further in the Discussion section. Figure 6(b) shows examples of two tracks identified in ERAI that do not match with IBTrACS: the track labeled ”Storm 1” occurs in January of 2011 and is a storm that possibly occurs in the RMSC Nadi advisories, named 02F, but is not in IBTrACS, probably because it did not develop further into a true TS. Even so, it seems a substantial storm with 10m winds in ERAI over 20 m s^{-1} whilst near Australia. Figure 6(c) shows the infrared satellite image, which presents an asymmetric structure, unlike a true TS, with this storm more likely to be a hybrid warm core TC. The second storm shown in Figure 6(b) originates south of 20S, where very few IBTrACS storms have their genesis. This particular storm seems to have formed in the vicinity of the South Pacific Convergence Zone (SPCZ) and travels south eastward with relatively weak 10m winds in ERAI $\sim$15 m s^{-1} through a region of very little habitable land. It has no reference in any tropical storm advisories, yet its structure in the satellite imagery (Figure 6(d) shows some similarities with ”Storm 1” (Figure 6(c)) and it may also be a hybrid TC. As shown by Yanase et al. (2014) (Figure 1) using the Hart phase space classification of cyclones (Hart 2003), applied to reanalysis data, storms found between 20-40S in the SH summer tend to be hybrid storms. There are also storms in IBTrACS that do not match with an analysis track, but these tend to be the weakest storms below Category 1 as shown in Tables 3 and 4. These issues are further discussed in the Discussion section.

4. Discussion

There are several possibilities for the poorer performance of the objective detection method in the SH compared with the NH in terms of the detection, relative to the observed TCs in IBTrACS. As shown above, the discrepancy in numbers is closely associated with the weakest storms, tropical depressions and tropical storms (below Category 1). The first possibility for the differences between the NH and SH objective detection may be due to different biases in the best track data in the SH compared with the NH; the second is due to different biases in the representation of TCs in the reanalyses between the NH and SH; the third is due to the selection criteria used by the objective detection method to identify TCs in the reanalyses being not selective enough, or being mainly tuned to the NH. These will be addressed in turn.

In terms of possible biases in the IBTrACS observations, it is possible that the SH is observed differently than in the NH. The SH is sparsely inhabited in particular regions, such as the SP and SA, so that less emphasis may be placed on detection except for the most intense systems likely to make landfall (Kucas et al. 2014). Related to this is the application of different storm detection procedures in the different warning centers that produce the best track data (Velden and Coauthors 2006b; Kueh 2012). Storm classification is primarily based on the interpretation of satellite observations using empirical relationships such as the Dvorak scheme (Velden and Coauthors 2006a); there is little aircraft reconnaissance apart for the North Atlantic with some other limited coverage associated with field campaigns and in specific regions, e.g. Taiwan (DOTSTAR) (Wu and Coauthors 2005). The uncertainties of applying operational detection and classification schemes when storms are relatively weak and show a poor organization (Torn and Snyder 2012) may make deciding between whether a tropical disturbance should be classified as a tropical depression and counted in best track, or is some other tropical storm such as a subtropical or hybrid cyclone, difficult and dependent on subjective forecaster interpretation. Gyakum (2011) states that ”there is presently no single set of objective criteria that, if applied operationally, would irrefutably support a forecasters analysis of cyclone type (subtropical, hybrid or tropical)”. It is also unclear whether all agencies report weaker storms such as tropical depressions consistently in their best track analyses, and hence whether they make their way into IBTrACS. For example, HURDAT, produced by the National Hurricane Center (NHC), and which forms part of IBTrACS, and covers the North Atlantic and North Eastern Pacific includes subtropical cyclones (Landsea and Franklin 2013) where as the Joint Typhoon Warning Center (JWTC), which covers the Western North Pacific, South Pacific and South and North Indian Oceans, do not routinely include subtropical cyclones (Kucas et al. 2014; Gyakum 2011) unless they undergo Tropical Transition (TT) (Bentley et al. 2016; McTaggart-Cowan et al. 2013). Even within a single ocean basin where multiple agencies are operational, considerable uncertainties exist between different best track data sets. For example, Ren et al. (2011) and Barcikowska et al. (2012), highlight significant differences between JTWC and Japan Meteorological Agency (JMA) best track data in the Western North Pacific (WNP) in terms of frequency and intensity of TCs, with better agreement for frequencies for Category 2 TS and above; this is exactly where our objective detection scheme performs best in both hemispheres.

Therefore, uncertainties in the interpretation of the observations for the weaker tropical storms, and different agency operational procedures, may result in their exclusion from the best track archive. Several reassessments of best track data, in particular in the SH, have resulted in the inclusion of some additional storms, but also the removal of some others (Diamond et al. 2012) so that actual numbers are not significantly changed. However, evidence that the SH may be being treated differently for tropical storms in the observations than in the NH, in particular with respect to the weaker sub-tropical and hybrid storms, can be seen by considering the tropical storm advisories. Information on weak tropical disturbances, together with TCs, is available in text based reports from the warning agencies, such as the JTWC ”significant tropical cyclone advisories”. However, not all this information is included in the best track post season analysis and hence IBTrACS. For example, in the South Pacific, IBTrACS reports 5 storms in the 2011/2012 season (July-June) but scanning the advisories (Regional Specialized Meteorological Center (RSMC) Nadi) results in a much larger number of tropical disturbances, $\sim$20. A more quantitative comparison can be made using the combined advisories from each warning center, for each year, in each hemisphere (July-June in the SH). This information has been collated by Padgett and Young (2016) from 1998 onwards for both hemispheres, although some very weak systems are not included. Comparing the numbers in the advisories with those in IBTrACS over the period 1998-2012, which overlaps with our study period; in the NH, IBTrACS has on average 69 storms per year and the advisories 72, hence the advisories have $\sim$4% more storms, whereas for the SH, IBTrACS has on average 28 storms per year and the advisories 39, hence the advisories have $\sim$40% more storms. Hence in the NH it appears that a much larger proportion of the storms in the advisories make their way into the best track data than in the SH. This can partially explain the difference in numbers between IBTrACS and the TCs identified by the objective detection method in the reanalyses in the SH. It was discussed in the Matching Against IBTrACS subsection that some of the storms identified in the reanalyses appeared to be in the advisories but not IBTrACS.

Tropical disturbances and subtropical cyclones occur in all the ocean basins, and it seems that whether or not they contribute to the best track data may vary between the NH and SH and be dependent on the warning center procedures. The SPCZ and South Atlantic Convergence Zone (SACZ)are known to be associated with weak tropical depressions and subtropical cyclones in the

SH, as well as more intense tropical cyclones in the South Pacific (Vincent et al. 2011). A similar situation occurs in the North Pacific associated with the Mei-Yu front (Lee et al. 2006). The South Atlantic is not known as a very active TC region, due to relatively cool Sea Surface Temperatures and relatively high vertical wind shear. However, several studies have highlighted this region as susceptible to the formation of subtropical cyclones (Evans and Braun 2012; Gozzo et al. 2014) often in association with the SACZ. This is also seen in simulations produced with high-resolution GCMs where they are often identified as TCs (Roberts et al. 2015). The study of Gozzo et al. (2014), based on reanalysis data, found on average 7 subtropical cyclones per year with genesis between 20-30^{0}S, a number that is remarkably similar to the number of systems objectively detected in the reanalyses in this study in the SA region. The majority of the sub-tropical cyclones identified by Gozzo et al. (2014) do not seem to have made it into the advisories or best track data, either because they are to weak, even for the advisories, or possibly because in general they are moving away from land and therefore not a threat (Kucas et al. 2014). Another possibility is that SA sub-tropical cyclones are more asymmetric than those found in the North Atlantic (Evans and Braun 2012) and hence do not satisfy the criteria for inclusion in the TC best tracks. A similar situation may also occur in the South Pacific. If these additional uncertainties in the best track data are considered together with the numbers in the advisory data, then the actual numbers of TCs occurring in the SH may not be too far away from the numbers objectively identified here in the reanalyses. The results from the Matching Against IBTrACS subsection suggest that some of the differences between numbers in the SH between the objective identification used in this study and IBTrACS may be related to the identification of hybrid or sub-tropical cyclones by the objective identification scheme.

Other regions where subtropical or hybrid storms may need to be considered are the cool seasons in the eastern north Pacific, where they are called Kona storms (Kodama and Businger 1998).

Monsoon depressions may also be confused with weak tropical cyclones in the reanalyses as these also have a warm core aloft structure and occur in the north and south Indian ocean, western Pacific and Australian region (Hurley and Boos 2015). They represent an additional uncertainty in the best track archive, as they are occasionally included in the best track data in the western Pacific via the JTWC (Hurley and Boos 2015); however, as with subtropical cyclones, this is not done consistently for all agencies. These may also contribute to uncertainties in the best track data in the north and south Indian ocean and South Pacific.

The second possibility for the differences in the numbers of TCs detected by the objective detection method in the reanalyses and IBTrACS in the NH and SH concerns the quality of the reanalyses in the two hemispheres, which may affect how TCs are represented and hence contribute to the uncertainties in their detection in the reanalyses. The primary observations assimilated in the SH come from satellite observing platforms, which generally provide relatively coarse vertical resolutions, whereas in the NH the surface-based observing system provides a more diverse range of observations, including from sondes and aircraft. The use of direct satellite radiance assimilation, variational bias correction and modern assimilation methods has resulted in much better extraction of the information content in the observations, including for older observations (Rienecker et al. 2012). Discriminating between weak TSs, sub-tropical cyclones and other systems in the reanalyses is a problem in both hemispheres for the objective detection method, but could be more of a problem in the SH if the TCs are not as well simulated and storms, including sub-tropical or hybrid storms, do not have the correct structure. This could be exacerbated if there are more of the weaker type of storms in the SH associated with the convergence zones as discussed above, which, allied to the difficulty in separating these storms from other systems, may be a factor in the differences between the number of storms in IBTrACS and the number detected by the objective detection method in the reanalyses in the SH. The only way to test this is by using observing system experiments, where the NH observing system is degraded to that of the SH and the data assimilation re-run. These types of experiments have been performed in the past and have shown the relative importance of the different types of observations used in the reanalyses and how changes to the observing system may affect the reanalysis (Bengtsson et al. 2004b; Whitaker et al. 2009). However, it is very time consuming and expensive to re-run modern data assimilation systems, even if we had access to the same systems used to produce the reanlyses used here. Hence this is beyond the scope of this paper. However, studies using the same detection criteria as used here, applied to relatively high resolution climate model simulations for the current climate (Gleixner et al. 2013; Strachan et al. 2013; Roberts et al. 2015), have found similar results to those found here for the reanalyses, in that similar TC numbers to observations are found in the NH, albeit with some model dependent basin by basin biases, and a larger number of TCs than in the observations in the SH. This may indicate that the difference in the number of SH storms from the observations are not necessarily related to differences in the quality of the reanalyses in the two hemispheres, but more with possible biases in the best track data and possibly the detection criteria used in our objective scheme, discussed next.

The larger bias in the number of TCs identified by the objective detection method in the SH compared with the NH relative to observations may also be related to the detection criteria used here, and whether they are selective enough for the data used, so that more tropical depressions, subtropical cyclones and hybrid cyclones are identified as TCs, possibly related to the quality of the reanalyses as discussed above. TC detection schemes, applied to model or reanalysis data, are certainly sensitive to the detection criteria and tracking methodology employed (Horn et al. 2014), especially for weaker storms, as shown in this paper, and are most often tuned for the NH. An alternative approach would be to apply more selective criteria to remove subtropical and hybrid cyclones from the detection, based on previous studies focussed on studying subtropical cyclones, for example the Hart phase space parameters (Guishard et al. 2009; Evans and Braun 2012; Yanase et al. 2014). Another idea in the literature suggests using TC development pathways (McTaggart-Cowan et al. 2013), whereby tropical cyclogenesis is categorized according to dynamical metrics, although this would necessarily introduce added complexity and possibly more parameters to choose subjectively. It would also remove these types of storms in the NH, so that, whilst the numbers detected in the SH may compare better with the observations, the numbers may compare less favourably in the NH. However, it might allow a better focus on the different storm types.

It is likely that all three of the issues discussed above can lead to TC detection biases in the reanalyses relative to the best track data.

No TC tracking and/or identification scheme will be perfect and, although TC identification schemes can be re-tuned against the observations separately for the NH and SH or for individual ocean basins if necessary (Camargo and Zebiak 2002) to take account of possible deficiencies in the detection and the observational biases, this does not seem like a good idea if TC detection is to be applied to model simulations where methodological consistency is important.

5. Summary and conclusions

The study of TCs in six recent reanalyses has shown that all the reanalyses are capable of representing nearly all the TCs present in the best track archive of IBTrACS, with a detection rate of $\sim 98\%$ in the period since 2000 and slightly lower before this. However, how well the TCs are represented in the reanalyses, in terms of their properties, is less encouraging, with wind intensities significantly lower than in the observations and pressures too high in value. Although significant amounts of observations are assimilated by the data assimilation systems used in the reanalyses, in particular from satellites, this is unable to correct these deficiencies in the TC properties, due to the still too low model resolution and dependence on parameterized processes used in the reanalyses. Additional methods of assimilating observations in the vicinity of the TCs and vortex relocation can help improve this situation, but not to the extent where intensities get anywhere near those observed at current reanalysis resolutions. However, it is apparent that there has been some improvements in the representation of TCs in the more recent reanalyses of NCEP-CFSR, JRA55 and MERRA2; in particular MERRA2 shows a significant improvement over the older MERRA reanalysis in terms of wind and MSLP intensities. Separation distances between TCs identified in the reanalyses and the observations have also improved with the more recent reanalyses. The improvements in the intensities and location are most likely due to the increases in model horizontal resolutions and the use of improved data assimilation and bias correction systems, which are capable of extracting more information content from the older observations, as well as resulting in less observation rejection and the introduction of new and better calibrated observing systems in recent years. This progress is likely to continue as new reanalyses are produced with ever higher resolutions, such as the new ECMWF ERA5 reanalysis. Further improvements in data assimilation are expected as well as the introduction of new and more accurate observing systems, although the downside to this may be the introduction of spurious trends in TC properties.

The other aspect explored in this study is how well objective TC detection schemes are capable of detecting the same TCs that are in the observations using a widely used identification scheme. This is important in order to have confidence in these schemes when applied to climate model simulations and for comparisons made between models or experiment scenarios. This part of the study highlighted the problem of detecting TCs at the low intensity end of the TC intensity range: in particular, tropical depressions and up to category 1 (Saffir-Simpson), with gradual improvements in the detection rate with increasing TS category. This raises several issues: are the current detection schemes used at operational centers and for climate studies of TCs, which all have a rather similar methodology of user chosen thresholds on intensity and/or structure, selective enough; are TCs represented well enough in the reanalyses; are there problems with observational biases for weak storms? The answer to these questions is probably that all three play a role in differences found between the objective identification of TCs in reanalyses and the observed best track data. It is clear the intensities, and probably structure, are not well enough simulated in the reanalyses, which will cause problems when trying to discriminate between weak TSs and other tropical systems. In terms of more selective criteria, other approaches could certainly be introduced, such as the phase space approach, but this will also depend on how well TCs are represented in the reanalyses and the introduction of subjective thresholds on the phase space parameters (Yanase et al. 2014). However, it may be useful in removing the need for artificial boundaries in the TC identification such as the latitude band for genesis used in this study. The problem of observational bias is also an important aspect, in particular for the weaker storms, since forecaster interpretation and subjectivity will play a role in whether a particular storm is included in the best track data, as not all storms fall neatly into particular classifications. Allied to this are the different operational criteria employed by the different RSMC, which contribute data to the best track archives, such as whether to include tropical depressions or subtropical cyclones. This is likely the primary cause of the differences between the number of TCs identified in the reanalyses and IBTrACS, in particular in the SH. This makes the observations less than ideal for calibrating TC identification and tracking schemes, or indeed in their use in global climatological studies of TC frequencies and variability. It could be concluded that, given the uncertainties in the best track data sets, they should not be considered climate quality data sets. Better coordination between the RSMCs would help this situation going forward, although this is not necessarily part of their remit and their operational procedures are tailored to their region of responsibility. The problems of objectively classifying TCs operationally has been recognized by the Seventh International Workshop on Tropical Cyclones who have suggested that "a substantial contribution to the operational TC forecasting community could be made by recommending a universal cyclone classification methodology based on the latest research, operational forecasting capabilities, and real-time data availability". A re-evaluation of the observational record over the satellite period using a combination of the satellite data and reanalyses, using consistent identification methods for all basins, could perhaps resolve the observational bias problem over historical periods covered by the satellites and provide a more complete record of tropical storms for use in risk assessment and validating climate models. There has been some discussion that tropical depressions and subtropical cyclones should be included in the best track data for consistency (McAdie et al. 2009), since, before satellite observations became available, some subtropical systems were probably classified as TSs. Tropical depressions and subtropical cyclones are also associated with severe weather with TS like properties of strong winds and precipitation (Guishard et al. 2009; Gyakum 2011), so their inclusion can be justified in terms of their impact and for a more complete record of TC activity.

Whilst there are deficiencies in the representation of TCs in the reanalyses, and 10 m winds in particular should be used with caution, they can be complementary to the observations and provide added value information on TCs such as the pre- and post-TC stages of the lifecycle. For example, the tracking method used here identifies these earlier and later lifecycle stages, which can then be used to study the early development of TCs and their environment and the extratropical transition (Studholme et al. 2015) and how storms behave after this. The extratropical transition and its aftermath are becoming increasingly important for risk analysis at high latitudes following cases such as hurricane Sandy and Gonzalo and recent studies such as Haarsma et al. (2013) and is a known contributor to forecast uncertainty in the extra-tropics (Anwender and Harr 2008).

Acknowledgments. The authors would like to thank the various data archive centers for making the data used in this study available. ERA-Interim data provided courtesy ECMWF. MERRA was developed by the Global Modeling and Assimilation Office and supported by the NASA Modeling, Analysis and Prediction Program. Source data files can be acquired from the Goddard Earth Science Data Information Services Center (GES DISC). The NCEP-CFSR data set used for this study is provided from the Climate Forecast System Reanalysis (CFSR) project carried out by the Environmental Modeling Center (EMC), National Centers for Environmental Prediction (NCEP). The JRA25 data set used for this study is provided from the Japanese 25-year Reanalysis (JRA-25), a cooperative research project carried out by the Japan Meteorological Agency (JMA) and the Central Research Institute of Electric Power Industry (CRIEPI). The JRA55 data set used for this study is provided from the Japanese 55-year Reanalysis (JRA-55) project carried out by the Japan Meteorological Agency (JMA). Vidale acknowledges funding from the Willis Chair in Climate System Science and Climate Hazards. Cobb acknowledges funding from the "Innovate UK" Knowledge Transfer Partnership (KTP). The authors would also like to thank the two reviewers for their constructive comments leading to an improved paper.

References

Adeola, F. O., and J. S. Picou, 2014: Social capital and the mental health impacts of hurricane katrina: assessing long-term patterns of psychosocial distress. *International Journal of Mass Emergencies and Disasters*, **32 (1)**, 121–156.

Anwender, D., and S. C. Harr, P. A .and Jones, 2008: Predictability associated with the downstream impacts of the extratropical transition of tropical cyclones: Case studies. *Mon. Wea. Rev*, **136**, 3226–3247.

Barcikowska, M., F. Feser, and H. Von Storch, 2012: Usability of best track data in climate statistics in the western north pacific. *Monthly Weather Review*, **140 (9)**, 2818–2830.

Bell, R., J. Strachan, P. L. Vidale, K. I. Hodges, and M. Roberts, 2013: Response of tropical cyclones to idealized climate change experiments in a global high-resolution coupled general circulation model. *J. Clim.*, **26**, 7966–7980.

Bengtsson, L., M. Botzet, and M. Esch, 1995: Hurricane-type vortices in a general circulation model. *Tellus*, **47A**, 175196.

Bengtsson, L., S. Hagemann, and K. I. Hodges, 2004a: Can climate trends be calculated from reanalysis data?jgr-atmos,. *J. Geophys. Res. Atmos.*, **109**, doi:D11111,doi:10.1029/2004JD004536.

Bengtsson, L., K. Hodges, and M. Esch, 2007a: Tropical cyclones in a t159 resolution global climate model: Comparison with observations and re-analyses. *Tellus A*, **59 (4)**, 396–416.

Bengtsson, L., K. I. Hodges, M. Esch, N. Keenlyside, L. Kornblueh, J.-J. Luo, and T. Yamagata, 2007b: How may tropical cyclones change in a warmer climate? *Tellus*, **59A**, 539–561.

Bengtsson, L., K. I. Hodges, and S. Hagemann, 2004b: Sensitivity of the era40 reanalysis to the observing system: determination of the global atmospheric circulation from reduced observations. *Tellus*, **56A**, 456–471.

Bentley, A., M. D. Keyser, and L. F. Bosart, 2016: A dynamically based climatology of subtropical cyclones that undergo tropical transition in the north atlantic basin. *Mon. Wea. Rev*, **144**, 2049–2068.

Bloom, S. C., L. L. Takacs, A. M. da Silva, and D. Ledvina, 1996: Data assimilation using incremental analysis updates. *Mon. Wea. Rev.*, **124**, 12561271.

Bosilovich, M. G., and coauthors, 2015: Merra-2: Initial evaluation of the climate. Technical Report Series on Global Modeling and Data Assimilation, Volume 43 NASA/TM2015-104606/Vol. 43, National Aeronautics and Space Administration, Goddard Space Flight Center, Greenbelt, Maryland 20771, USA.

Camargo, S. J., A. G. Barnston, , and S. E. Zebiak, 2005: A statistical assessment of tropical cyclone activity in atmospheric general circulation models. *Tellus*, **57A**, 589604.

Camargo, S. J., and . Zebiak, S. E. 17, 2002: Improving the detection and tracking of tropical storms in atmospheric general circulation models. *Wea. Forecasting*, **17**, 11521162.

Davis, C., and Coauthors, 2008: Prediction of landfalling hurricanes with the advanced hurricane wrf model. *Mon. Wea. Rev.*, **136**, 19902005.

Dee, D., and Coauthors, 2011: The era-interim reanalysis: Configuration and performance of the data assimilation system. *Quarterly Journal of the Royal Meteorological Society*, **137 (656)**, 553–597.

Dee, D. P., and S. M. Uppala, 2009: Variational bias correction of satellite radiance data in the era-interim reanalysis. *Q. J. R. Meteorol. Soc.*, **135**, 18301841.

Diamond, H., A. M. Lorrey, K. R. Knapp, and D. H. Levinson, 2012: Development of an enhanced tropical cyclone tracks database for the southwest pacific from 1840-2010. *Int. J. Climatol.*, **32**, 22402250.

ECMWF, 2015: *Part IV: Physical Processes*. IFS Documentation, ECMWF, operational implementation 12 May 2015.

Evans, J. L., and A. Braun, 2012: A climatology of subtropical cyclones in the south atlantic. *J Climate*, **25**, 73287340.

Fine, C. M., R. H. Johnson, P. E. Ciesielski, and R. K. Taft, 2016: The role of topographically induced vortices in tropical cyclone formation over the indian ocean. *Mon. Wea. Rev*, **144**, 4827–4847.

for the Caribbean, S. H., 2009: Belize: Macro socio-economic assessment of the damage and losses caused by tropical depression 16. Tech. rep., United Nations Economic Commission for Latin America and the Caribbean.

Gleixner, S., N. Keenlyside, K. I. Hodges, W.-L. Tseng, and L. Bengtsson, 2013: An inter-hemispheric comparison of the tropical storm response to global warming. *Climate Dynamics*, **42 (7-8)**, 2147–2157, doi:10.1007/s00382-013-1914-6, URL http://dx.doi.org/10.1007/s00382-013-1914-6.

Gozzo, L. F., R. P. da Rocha, M. S. Reboita, and S. Sugahara, 2014: Subtropical cyclones over the southwestern south atlantic: Climatological aspects and case study. *J. Climate*, **27**, 85438562.

Guishard, M. P., J. L. Evans, and R. E. Hart, 2009: Atlantic subtropical storms. part ii: Climatology. *J. Climate*, **22**, 35743594.

Gyakum, J. R., 2011: Subtropical and hybrid systems. *Proc. Seventh Int. Workshop on Tropical Cyclones*, Saint-Gilles-Les-Bains, La Runion, France, World Meteorological Organization, 1.6.1-1.6.28, WMO/TD-1561.

Haarsma, R. J., W. Hazeleger, C. Severijns, H. de Vries, A. Sterl, R. Bintanja, G. J. van Oldenborgh, and H. W. van den Brink, 2013: More hurricanes to hit western europe due to global warming. *Geophys. Res. Lett.*, **40**, 17831788.

Harper, B., J. Kepert, and J. Ginger, 2010: Guidelines for converting between various wind averaging periods in tropical cyclone conditions. *World Meteorological Organization, WMO/TD*, **1555**.

Hart, R. E., 2003: A cyclone phase space derived from thermal wind and thermal asymmetry. *Monthly Weather Review*, **131 (4)**, 585–616.

Hatsushika, H., J. Tsutsui, M. Fiorino, and K. Onogi, 2006: Impact of wind profile retrievals on the analysis of tropical cyclones in the jra-25 reanalysis. *ł. 2* , **84 (5)**, 891–905.

Hodges, K. I., 1994: A general-method for tracking analysis and its application to meteorological data. *Monthly Weather Review*, **122 (11)**, 2573–2586.

Hodges, K. I., 1995: Feature tracking on the unit sphere. *Monthly Weather Review*, **123 (12)**, 3458–3465.

Hodges, K. I., 1999: Adaptive constraints for feature tracking. *Monthly Weather Review*, **127 (6)**, 1362–1373.

Hodges, K. I., and R. Emerton, 2015: The prediction of northern hemisphere tropical cyclone extended life cycles by the ecmwf ensemble and deterministic prediction systems. part i: Tropical cyclone stage. *Monthly Weather Review*, **143**, 50915114, doi:10.1175/MWR-D-13-00385.1.

Hodges, K. I., B. J. Hoskins, J. Boyle, and C. Thorncroft, 2003: A comparison of recent reanalysis datasets using objective feature tracking: Storm tracks and tropical easterly waves. *Monthly Weather Review*, **131 (9)**, 2012–2037.

Horn, M., and Coauthors, 2014: Tracking scheme dependence of simulated tropical cyclone response to idealized climate simulations. *Journal of Climate*, **27 (24)**, 9197–9213.

Hurley, J. V., and W. R. Boos, 2015: A global climatology of monsoon low-pressure systems. *Q.J.R. Meteorol. Soc.*, **141**, 10491064.

Jones, S. C., and Coauthors, 2003: The extratropical transition of tropical cyclones: Forecast challenges, current understanding, and future directions. *Weather and Forecasting*, **18**, 1052–1092.

Kilic, C., and C. C. Raible, 2013: Investigating the sensitivity of hurricane intensity and trajectory to sea surface temperatures using the regional model wrf. *Meteorologische Zeitschrif*, **22**, 685698.

Kimball, S. K., and M. S. Mulekar, 2004: A 15-year climatology of north atlantic tropical cyclones. part i: Size parameters. *J. Clim.*, **17**, 3555–3575.

King, R. O., 2013: Financing natural catastrophe exposure: Issues and options for improving risk transfer markets. ...., Congressional Research Service, Library of Congress.

Knaff, J. A., and R. M. Zehr, 2007: Reexamination of tropical cyclone windpressure relationships. *Wea. Forecasting*, **22**, 7188.

Knapp, K. R., M. C. Kruk, D. H. Levinson, H. J. Diamond, and C. J. Neumann, 2010: The international best track archive for climate stewardship (ibtracs) unifying tropical cyclone data. *Bulletin of the American Meteorological Society*, **91 (3)**, 363–376.

Kobayashi, S., and Coauthors, 2015: The jra-55 reanalysis: General specifications and basic characteristics. *J. Met. Soc. Jap.*, **93**, 5–48, doi:10.1175/JCLI-D-14-00005.1.

Kodama, K., and S. Businger, 1998: Weather and forecasting challenges in the pacific region of the national weather service. *Wea. Forecasting*, **13**, 253276.

Kucas, M., S. J. Barlow, and R. C. Ballucanag, 2014: Subtropical cyclones: Operational practices and analysis methods applied at the joint typhoon warning center. *Extended abstracts, 31st Conference on Hurricanes and Tropical Meteorology*, San Diego, CA, Amer. Meteor. Soc., 1-8.

Kueh, M.-T., 2012: Multiformity of the tropical cyclone windpressure relationship in the western north pacific: discrepancies among four best-track archives. *Environ. Res. Lett.*, **7**, doi:10.1088/1748-9326/7/2/024015.

Landsea, C., 2007: Counting atlantic tropical cyclones back to 1900. *Eos, Transactions American Geophysical Union*, **88 (18)**, 197–202.

Landsea, C. W., and J. L. Franklin, 2013: Atlantic hurricane database uncertainty and presentation of a new database format. *Monthly Weather Review*, **141 (10)**, 3576–3592.

Landsea, C. W., G. A. Vecchi, L. Bengtsson, and T. R. Knutson, 2009: Impact of duration thresholds on atlantic tropical cyclone counts. *J. Climate*, **23**, 25082519, doi:10.1175/2009JCLI3034.1.

Lee, C.-S., Y.-L. Lin, and K. K. Cheung, 2006: Tropical cyclone formations in the south china sea associated with the mei-yu front. *Mon. Wea. Rev.*, **134**, 26702687.

Manganello, J., and Coauthors, 2012: Tropical cyclone climatology in a 10-km global atmospheric gcm: Toward weather-resolving climate modeling. *J. Clim.*, **25,3867-3893.**, 3867–3893.

McAdie, C. J., C. W. Landsea, C. J. Neumann, J. E. David, E. S. Blake, and G. R. Hammer, 2009: *Tropical cyclones of the north atlantic ocean, 1851 2006*. pp237, National Climatic Data Center, Asheville, NC, in cooperation with the National Hurricane Center, Miami, FL.

McTaggart-Cowan, R., T. J. Galarneau, L. F. Bosart, R. W. Moore, and O. Martius, 2013: A global climatology of baroclinically influenced tropical cyclogenesis. *Mon. Wea. Rev.*, **141**, 1963–1989.

Molod, A., L. Takacs, M. Suarez, and J. Bacmeister, 2015: Development of the geos-5 atmospheric general circulation model: evolution from merra to merra2. *Geosci. Model Dev.*, **8**, 13391356.

Murakami, H., 2014: Tropical cyclones in reanalysis data sets. *Geophysical Research Letters*, **41 (6)**, 2133–2141.

Onogi, K., and Coauthors, 2007: The jra-25 reanalysis. *J. Met. Soc. Jap*, **85**, 369–432, doi:10.2151/jmsj.85.369.

Padgett, G., and S. Young, 2016: Global tropical cyclones summaries and operational track data. URL https://australiasevereweather.com/cyclones/tropical_cyclone_summaries_track_data.htm, Accessed 13/10/2016.

Ren, F., J. Liang, G. Wu, W. Dong, and X. Yang, 2011: Reliability analysis of climate change of tropical cyclone activity over the western north pacific. *J Climate*, **24**, 58875898.

Rienecker, M. M., and Coauthors, 2012: Atmospheric reanalyses recent progress and prospects for the future. a report from a technical workshop, april 2010. NASA Technical Report Series on Global Modeling and Data Assimilation, 29 NASA TM2012-104606.

Rienecker, M. R., and coauthors, 2011: Merra - nasa's modern-era retrospective analysis for research and applications. *J. Climate*, **24**, 3624–3648, doi:10.1175/JCLI-D-11-00015.1.

Roberts, M. J., and Coauthors, 2015: Tropical cyclones in the upscale ensemble of high-resolution global climate models*. *Journal of Climate*, **28 (2)**, 574–596.

Roth, D. M., 2002: A fifty year history of subtropical cyclones. *Preprints, 25th Conf. on Hurricanes and Tropical Meteorology, San Diego, CA*, Amer. Meteor. Soc., P1.43, URL http://ams.confex.com/ams/pdfpapers/37402.pdf.

Sardeshmukh, P. D., and B. J. Hoskins, 1984: Spectral smoothing on the sphere. *Mon. Wea. Rev.*, **112**, 2524–2529.

Schenkel, B. A., and R. E. Hart, 2012: An examination of tropical cyclone position, intensity, and intensity life cycle within atmospheric reanalysis datasets. *J. Climate*, **25**, 34533475.

Schreck III, C. J., K. R. Knapp, and J. P. Kossin, 2014: The impact of best track discrepancies on global tropical cyclone climatologies using ibtracs. *Monthly Weather Review*, **142 (10)**, 3881–3899.

Serra, Y. L., G. N. Kiladis, and K. I. Hodges, 2010: Tracking and mean structure of easterly waves over the intra-americas sea. *Journal of Climate*, **23 (18)**, 4823–4840.

Shaevitz, D. A., and Coauthors, 2014: Characteristics of tropical cyclones in high-resolution models in the present climate. *J. Adv. Model. Earth Syst.*, **6 , 11541172,**, 11541172,.

Shao, H., and Coauthors, 2016: Bridging research to operations transitions: Status and plans of community gsi. *Bulletin of the American Meteorological Society*, **97**, 1427–1440.

Strachan, J., P. L. Vidale, K. Hodges, M. Roberts, and M.-E. Demory, 2013: Investigating global tropical cyclone activity with a hierarchy of agcms: The role of model resolution. *Journal of Climate*, **26 (1)**, 133–152.

Studholme, J., K. I. K. I. Hodges, and C. M. Brierley, 2015: Objective determination of the extratropical transition of tropical cyclones in the northern hemisphere. *Tellus*, **67A**, 24 474, URL http://dx.doi.org/10.3402/tellusa.v67.24474.

Suranjana, S., and Coauthors, 2010: The ncep climate forecast system reanalysis. *Bulletin of the American Meteorological Society*, **91**, 1015–1057.

Torn, R. D., and C. Snyder, 2012: Uncertainty of tropical cyclone best-track information. *Weather and Forecasting*, **27 (3)**, 715–729.

Truchelut, R., R. Hart, and B. Luthman, 2013: Global identification of previously undetected pre-satellite era tropical cyclone candidates in noaa/cires 20th century reanalysis data. *J. Appl. Meteor. Climatol.*, **52**, 2243–2259, doi:10.1175/JAMC-D-12-0276.1.

Velden, C., and Coauthors, 2006a: The dvorak tropical cyclone intensity estimation technique: A satellite-based method that has endured for over 30 years. *Bulletin of the American Meteorological Society*, **87**, 1195–1210.

Velden, C., and Coauthors, 2006b: Supplement to: The dvorak tropical cyclone intensity estimation technique: A satellite-based method that has endured for over 30 years. *Bull.Amer.Met.Soc.*, **87**, S6–S9.

Vincent, E. M., M. Lengaigne, C. E. Menkes, N. C. Jourdain, P. Marchesiello, and . Madec, 2011: Interannual variability of the south pacific convergence zone and implications for tropical cyclone genesis. *Climate Dynamics*, **36**, 1881–1896.

Walsh, K. J. E., M. Fiorino, C. W. Landsea, and K. L. McInnes, 2007: Objectively determined resolution-dependent threshold criteria for the detection of tropical cyclones in climate models and reanalyses. *J. Clim.*, **20**, 23072314.

Whitaker, J. S., G. P. Compo, and J.-N. Thpaut, 2009: A comparison of variational and ensemble-based data assimilation systems for reanalysis of sparse observations. *Mon. Weather Rev.*, **137**, 19911999.

Wu, C.-C., and Coauthors, 2005: Dropwindsonde observations for typhoon surveillance near the taiwan region (dotstar). *Bull. Amer. Met. Soc.*, **86**, 787–790.

Yanase, W., H. Niino, K. I. Hodges, and N. Kitabatake, 2014: Parameter spaces of environmental fields responsible for cyclone development from tropics to extratropics. *J. Climate*, **27**, 652671.

LIST OF TABLES

TABLE 1. Summary of the reanalysis datasets used in this study. Abbreviations, 4D-Var, 4D variational assimilation ; 3D-Var, 3D variational assimilation; TL255L60, triangular truncation 255, linear grid, 60 vertical levels; GSI, Grid-point statistical interpolation; IAU, Incremental Analysis Update.

	$ERAI$	$JRA25$	$JRA55$	$NCEP-CFSR$	$MERRA$	$MERRA2$
Assimilation	4D-Var	3D-Var	4D-Var	3D-Var	3D-VAR	3D-Var
				GSI	GSI+IAU	GSI+IAU
Model	TL255L60	T106L40	TL319L60	T382L64	$1/2^0 \times 2/3^0$ L72	Cubed Sphere
Resolution	(80km)	(120km)	(55km)	(38km)	(55km)	(50km)
Data Grid	512×256	288×145	288×145	720×361	540×361	576×361

TABLE 2. The POD for the NH and SH for the direct matching method applied to the reanalysis tracks (c.f. Direct Matching Results section) and the POD and FAR for the NH and SH based on the objective detection method (c.f. Matching Against IBTrACS subsection).

	POD					
	ERAI	*JRA25*	*JRA55*	*NCEP−CFSR*	*MERRA*	*MERRA2*
NH Direct Match	0.95	0.95	0.95	0.95	0.95	0.95
NH Objective	0.60	0.76	0.80	0.70	0.51	0.67
SH Direct Match	0.93	0.93	0.94	0.93	0.90	0.93
SH Objective	0.76	0.84	0.87	0.83	0.61	0.79
	FAR					
NH Objective	0.28	0.16	0.29	0.36	0.21	0.36
SH Objective	0.60	0.43	0.58	0.58	0.54	0.63

TABLE 3. Storms that match and don't match with IBTrACS in the NH by storm category, for storms identified by the objective detection method applied to the reanalysis tracks, in the first column for each reanalysis, and for the direct matching method performed in the Direct Matching Results subsection in the second column with brackets for each reanalysis. Values are number per year.

Category		$ERAI$	$JRA25$	$JRA55$	$NCEP-CFSR$	$MERRA$	$MERRA2$
TD	Match	2.91 (7.94)	3.26 (7.94)	5.24 (8.03)	3.50 (8.00)	2.29 (7.85)	3.48 (7.67)
	No Match	5.56 (0.53)	5.21 (0.53)	3.24 (0.44)	4.97 (0.47)	6.18 (0.62)	4.91 (0.73)
TS	Match	11.85 (22.38)	18.62 (22.53)	18.32 (22.53)	14.76 (22.32)	9.85 (22.44)	14.24 (22.45)
	No Match	11.85 (1.32)	5.09 (1.18)	5.38 (1.18)	8.94 (1.38)	13.85 (1.26)	9.73 (1.52)
CAT1	Match	8.74 (12.23)	11.18 (12.23)	11.17 (12.24)	10.09 (12.12)	7.74 (12.21)	9.76 (12.33)
	No Match	3.44 (0.00)	1.00 (0.00)	1.00 (0.00)	2.09 (0.06)	4.44 (0.00)	2.55 (0.00)
CAT2	Match	5.29 (6.35)	6.15 (6.38)	6.00 (6.35)	5.82 (6.38)	4.76 (6.35)	5.64 (6.39)
	No Match	1.06 (0.00)	0.21 (0.00)	0.35 (0.00)	0.53 (0.00)	1.59 (0.00)	0.73 (0.00)
CAT3	Match	6.15 (7.00)	6.91 (7.06)	6.82 (7.03)	6.71 (7.06)	5.82 (7.03)	6.42 (7.06)
	No Match	0.88 (0.03)	0.12 (0.00)	0.21 (0.00)	0.32 (0.00)	1.21 (0.00)	0.64 (0.00)
CAT4	Match	5.97 (6.79)	6.76 (6.79)	6.71 (6.74)	6.47 (6.79)	5.76 (6.76)	6.48 (6.76)
	No Match	0.82 (0.00)	0.03 (0.00)	0.09 (0.06)	0.32 (0.00)	1.03 (0.03)	0.33 (0.06)
CAT5	Match	1.09 (1.12)	1.12 (1.12)	1.12 (1.12)	1.12 (1.12)	1.03 (1.12)	1.09 (1.09)
	No Match	0.03 (0.00)	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)	0.09 (0.00)	0.00 (0.00)

TABLE 4. Same as Table 2 but for the SH.

Category		*ERAI*	*JRA25*	*JRA55*	*NCEP − CFSR*	*MERRA*	*MERRA2*
TD	Match	0.42 (0.58)	0.48 (0.61)	0.52 (0.58)	0.48 (0.61)	0.21 (0.49)	0.44 (0.48)
	No Match	0.18 (0.03)	0.12 (0.00)	0.09 (0.03)	0.12 (0.00)	0.39 (0.12)	0.13 (0.06)
TS	Match	7.15 (9.09)	8.03 (9.06)	8.55 (9.09)	7.76 (9.18)	5.67 (9.00)	7.59 (8.91)
	No Match	2.42 (0.48)	1.55 (0.51)	1.03 (0.48)	1.82 (0.39)	3.91 (0.58)	2.00 (0.39)
CAT1	Match	4.55 (5.36)	5.09 (5.36)	5.12 (5.39)	4.94 (5.39)	3.79 (5.33)	4.75 (5.21)
	No Match	0.88 (0.06)	0.33 (0.06)	0.30 (0.03)	0.48 (0.03)	1.64 (0.09)	0.69 (0.06)
CAT2	Match	2.21 (2.64)	2.58 (2.61)	2.55 (2.64)	2.52 (2.64)	2.03 (2.61)	2.38 (2.61)
	No Match	0.61 (0.18)	0.24 (0.21)	0.27 (0.18)	0.30 (0.18)	0.79 (0.21)	0.53 (0.21)
CAT3	Match	2.55 (2.73)	2.61 (2.70)	2.70 (2.73)	2.64 (2.70)	2.12 (2.70)	2.63 (2.73)
	No Match	0.18 (0.00)	0.12 (0.03)	0.03 (0.00)	0.09 (0.03)	0.61 (0.03)	0.19 (0.00)
CAT4	Match	2.69 (2.76)	2.73 (2.73)	2.64 (2.76)	2.76 (2.76)	2.33 (2.76)	2.78 (2.76)
	No Match	0.06 (0.00)	0.03 (0.03)	0.12 (0.00)	0.00 (0.00)	0.42 (0.00)	0.06 (0.00)
CAT5	Match	0.58 (0.58)	0.51 (0.52)	0.55 (0.58)	0.55 (0.55)	0.55 (0.58)	0.59 (0.58)
	No Match	0.00 (0.00)	0.06 (0.06)	0.03 (0.00)	0.03 (0.03)	0.03 (0.00)	0.00 (0.00)

TABLE 5. The POD for the NH and SH for the TC obtained from the reanalyses by the objective detection method that match with the observed Cat1-Cat5 TS only.

	POD					
	$ERAI$	$JRA25$	$JRA55$	$NCEP-CFSR$	$MERRA$	$MERRA2$
NH Objective	0.81	0.96	0.95	0.90	0.75	0.87
SH Objective	0.88	0.91	0.95	0.94	0.75	0.92

LIST OF FIGURES

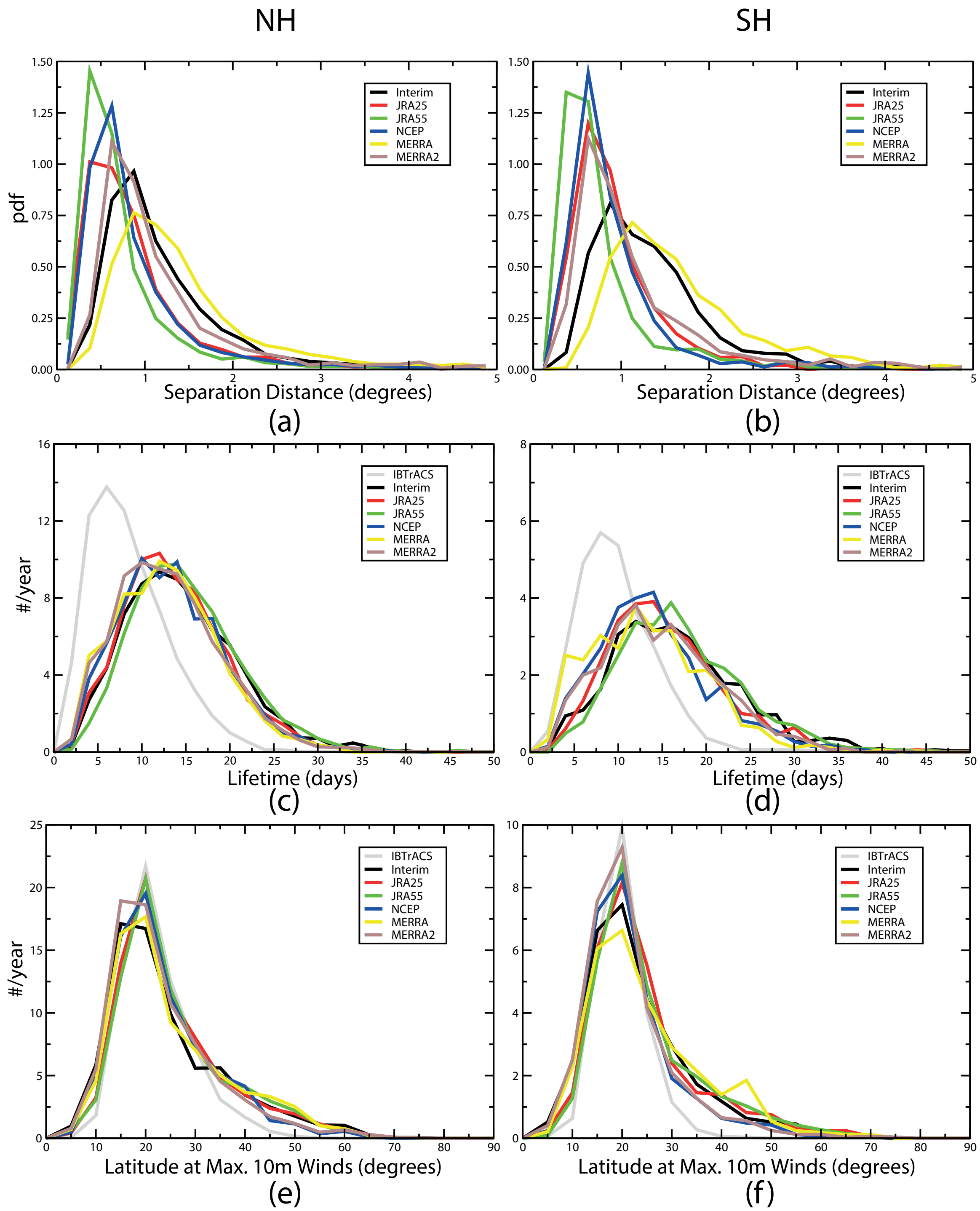

FIG. 1. Distribution of mean separation distances (geodesic degrees, $1^0 \simeq 111km$) between the reanalysis tracks and those of IBTrACS for tracks that match using the direct matching method (c.f. Direct Matching Results subsection) (a) NH, (b) SH, distribution of lifetimes (days) for the matched tracks (c) NH, (d) SH and the distribution of latitudes at which the matched tracks attain the peak intensity based on the 10m winds (e) NH, (f) SH.

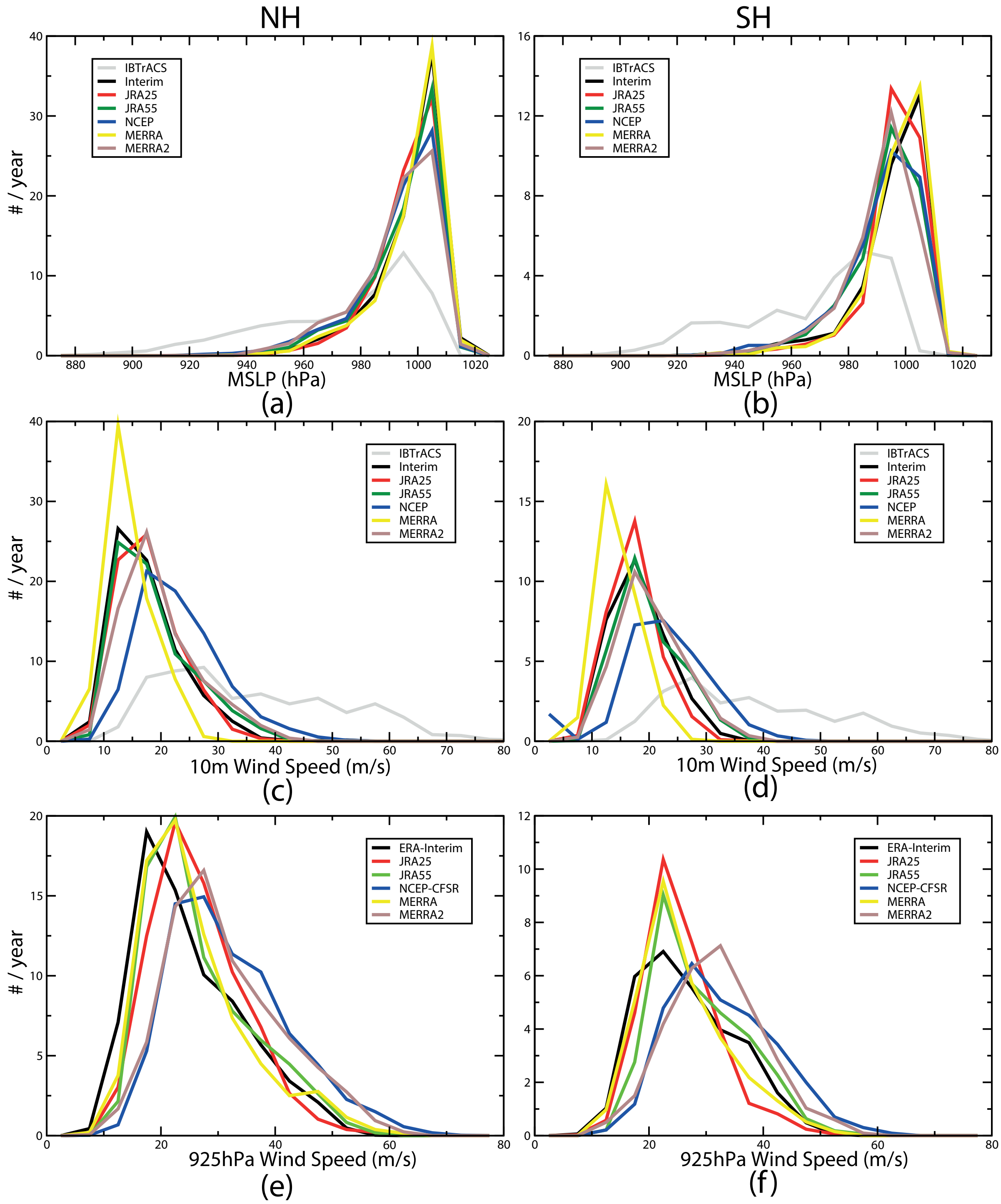

FIG. 2. Distributions for the peak attained intensities of matched reanalysis and IBTrACS tracks obtained using the direct matching method (c.f. Direct Matching Results subsection) based on the MSLP, 10m winds and 925hPa winds (not for IBTrACS), (a) NH MSLP (hPa), (b) SH MSLP (hPa), (c) NH 10m wind speed (m s^{-1}), (d) SH 10m wind speed (m s^{-1}), (e) NH 925hPa wind speed (m s^{-1}), (f) SH 925hPa wind speed (m s^{-1}).

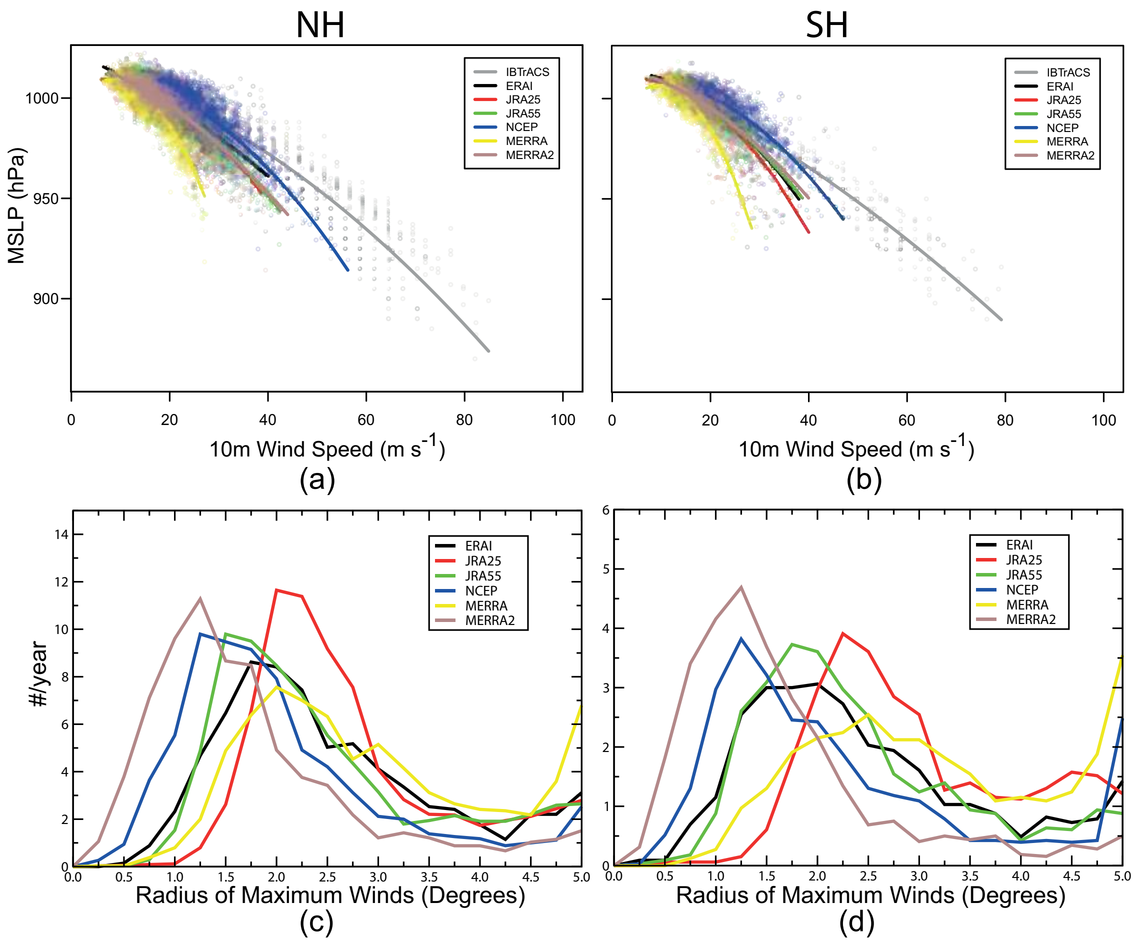

FIG. 3. Wind-pressure relationships for IBTrACS and each reanalysis and distributions for the radius of maximum winds for the reanalyses based on the direct matching method (c.f. Direct Matching Results subsection). (a) NH 10m wind speed versus MSLP, (b) SH 10m wind speed versus MSLP, (c) NH radius of maximum winds and (d) SH radius of maximum winds.

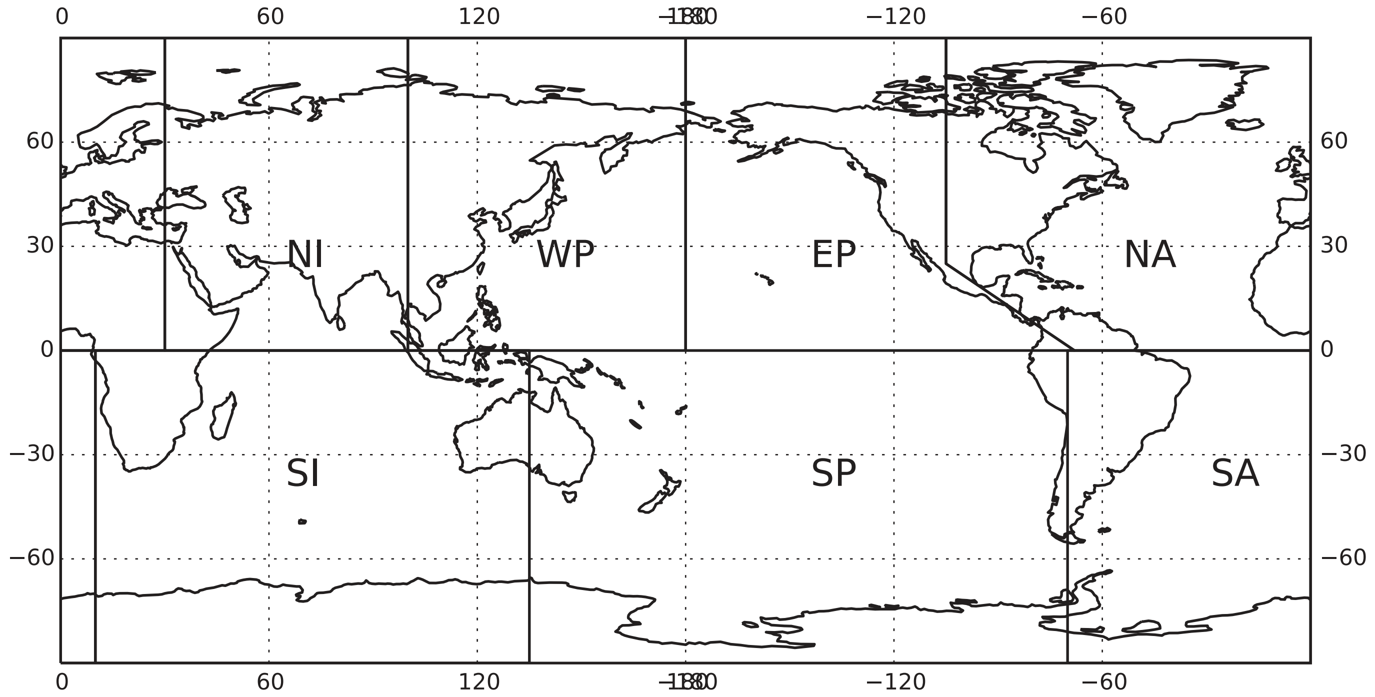

FIG. 4. The seven basins used in this study, based on the IBTrACS definition. NI: North Indian, WP: West Pacific, EP: East Pacific, NA: North Atlantic, SI: South Indian, SP: South Pacific, SA: South Atlantic.

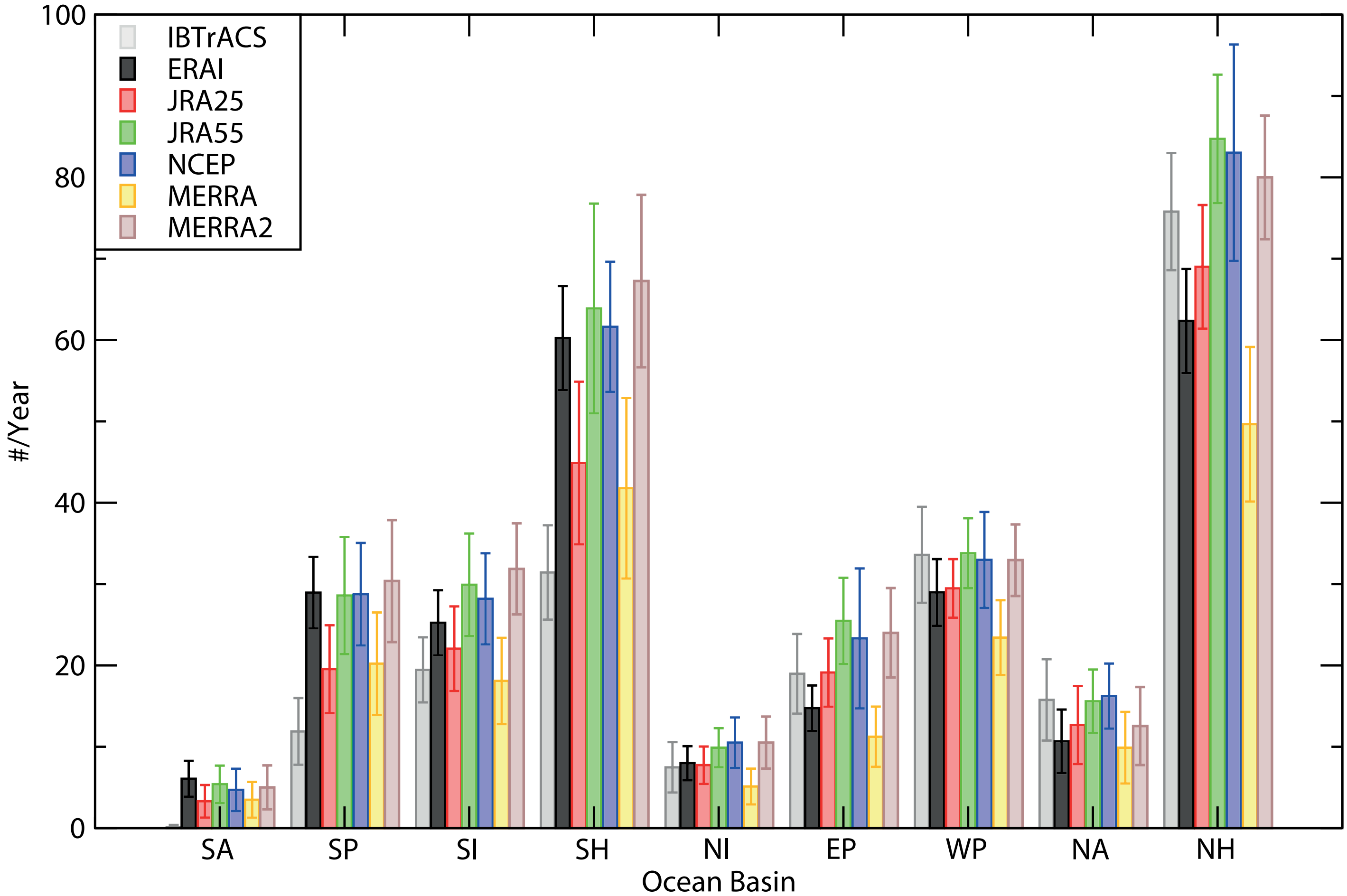

FIG. 5. The average number of TCs per year for each of the seven basins for IBTrACS and identified in the reanalyses based on the objective detection method (c.f. Objective Identification subsection). Vertical lines indicate the standard deviation. NI: North Indian, WP: West Pacific, EP: East Pacific, NA: North Atlantic, SI: South Indian, SP: South Pacific, SA: South Atlantic.

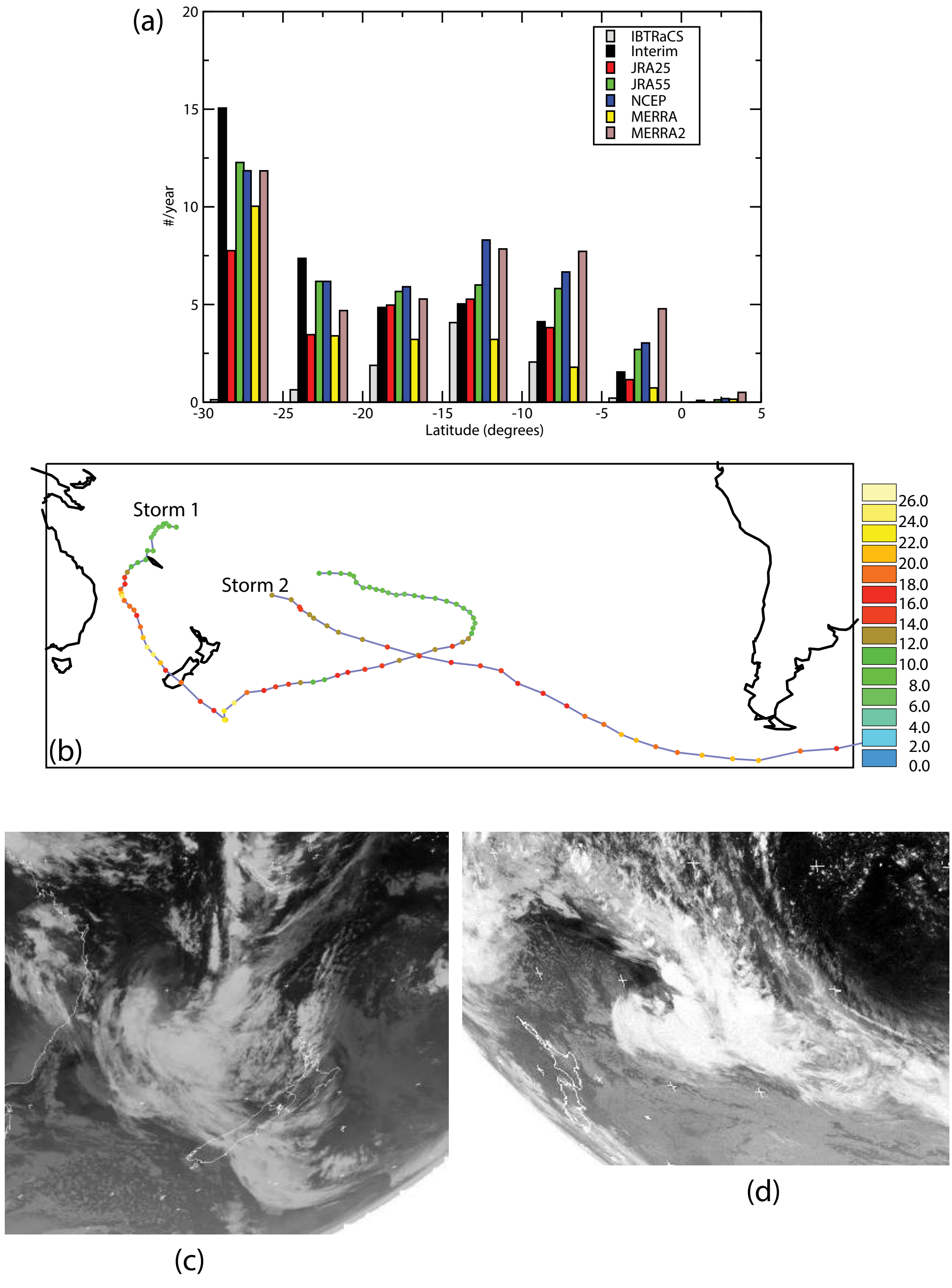

FIG. 6. (a) Latitude at which genesis occurs in the SH for the objectively identified TCs in the reanalyses that do not match with IBTrACS (number per year), (b) examples of two tracks identified in the ERAI reanalysis with no matching track in IBTrACS, coloured dots indicate 10m wind speeds (m s^{-1}), (c) MTSAT infrared satellite image of Storm 1 in (b) on 1800 UTC 1 Jan 2011 (d) GOES West infrared satellite image of Storm 2 in (b) on 1200 UTC 24 Dec 2011.