Advancing Tropical Cyclone Precipitation Forecast Verification Methods and Tools

Kathryn M. Newman[1,3], Barbara Brown[1,3], John Halley Gotway[1,3], Ligia Bernardet[2,3], Mrinal Biswas[1,3], Tara Jensen[1,3], and Louisa Nance[1,3]

[1]*National Center for Atmospheric Research, Boulder, CO*

[2]*Global Systems Laboratory, National Oceanic and Atmospheric Administration, Boulder, CO*

[3]*Developmental Testbed Center, Boulder, CO*

Corresponding author: Kathryn M. Newman, knewman@ucar.edu

1

ABSTRACT

Tropical cyclone (TC) forecast verification techniques have traditionally focused on track and intensity, as these are some of the most important characteristics of TCs and are often the principle verification concerns of operational forecast centers. However, there is a growing need to verify other aspects of TCs as process-based validation techniques may be increasingly necessary for further track and intensity forecast improvements as well as improving communication of the broad impacts of TCs including inland flooding from precipitation. Here we present a set of TC-focused verification methods available via the Model Evaluation Tools (MET) ranging from traditional approaches, to the application of storm-centric coordinates and the use of feature-based verification of spatially-defined TC objects.

Storm-relative verification using observed and forecast tracks can be useful for identifying model biases in precipitation accumulation in relation to the storm center. Using a storm-centric cylindrical coordinate system based on the radius of maximum wind adds additional storm-relative capabilities to regrid precipitation fields onto cylindrical or polar coordinates. This powerful process-based model diagnostic and verification technique provides a framework for improved understanding of feedbacks between forecast tracks, intensity, and precipitation distributions. Finally, object-based verification including land masking capabilities provides even more nuanced verification options. Precipitation objects of interest, either the central core of TCs or extended areas of rainfall after landfall, can be identified, matched to observations, and quickly aggregated to build meaningful spatial and summary verification statistics.

1. Introduction

Tropical cyclone (TC) forecast verification techniques have traditionally focused on track and intensity, as these are some of the most important characteristics of TCs and are often the principle verification concerns of operational forecast centers (Goerss 2000; Franklin et al. 2003; Sampson et al. 2008; Gall et al. 2013; NHC 2022). Additionally, TC genesis has also become integrated into the efforts of operational forecast centers (DeMaria et al. 2001; Halperin et al. 2013). As a result, many publications and verification tools have been developed for track, intensity, and genesis forecasts (Hemming 2017; Halperin et al. 2017; Brown et al. 2021; NHC 2022). Until recently, precipitation forecasts may not have been emphasized due to significant

2

track errors, poor representation of storm structure due to model resolution, and lack of availability of precipitation estimates, particularly over the ocean. However, there is a growing need to evaluate other aspects of TC predictions, with the understanding that additional verification metrics and process-based validation techniques may be increasingly necessary to enable continued improvements in track and intensity forecast skill (Kim et al. 2018; Cheung et al. 2018; Haiden et al. 2019), and improve communication of the broad impacts of TCs, including those associated with inland flooding from precipitation (Rappaport 2014; Morrow and Lazo 2015; Meléndez- Landaverde et al. 2020). Specifically, verification of TC precipitation can both inform improvements in process representation within models to improve storm evolution and help improve forecasts of the significant risks associated with extreme rainfall and flooding from landfalling TCs (Marchok et al. 2007; Cheung et al. 2018). Moreover, understanding the model-based quantitative precipitation forecasts (QPF) over the ocean is important, for example, because any forecast biases may impact the predicted storm characteristics and associated flooding potential at the time of landfall.

Over the past 20 years or more, several science advances have occurred that enable objective, large-sample verification of TC precipitation. First is the development of high-quality observed precipitation datasets with fine spatiotemporal resolution over both land and ocean. The current generation of US Weather Surveillance Radars (WSR-88Ds) was deployed throughout the 1990s with subsequent development of radar mosaics and gridded quantitative precipitation estimates (QPE) across the continental US. The Stage IV national rainfall QPE product began regular distribution in late 2001 (Lin and Mitchell 2005). At the same time, spaceborne QPE saw significant enhancements resulting from the Tropical Rainfall Measurement Mission (TRMM) satellite launched in 1997 and its associated QPE products becoming available in subsequent years (Haddad et al. 1997; Huffman 2006). New methods to better combine geostationary infrared and passive microwave precipitation estimates were also developed around the same time (e.g. Joyce et al. 2004). More recently, the National Aeronautics and Space Administration (NASA) Global Precipitation Measurement (GPM) mission has enabled development of the Integrated Multi-satellitE Retrievals for GPM (IMERG), which is an improved satellite precipitation product combining active, passive microwave, and geostationary satellite data (e.g. Huffman et al. 2020, Qi et al. 2021). Another important factor is the ever-increasing resolution

and skill of global and regional numerical weather prediction (NWP) models. Global NWP is now performed on ~10 km horizontal grids (Yang et al. 2020; ECMWF 2022) while regional NWP forecasts are often convective permitting with resolutions as small as 1-2 km (Clark et al. 2016; Biswas et al. 2018), with corresponding high spatio-temporal resolution QPF having been available for many years.

A limited number of studies have been undertaken to examine TC-focused QPF performance. Marchok et al. (2007) developed a method to verify TC QPF that focuses on large-scale patterns, mean and median precipitation, and extreme (95th percentile) precipitation, while also accounting for model-produced track errors to develop storm-relative verification statistics. Marchok et al. (2007) applied their technique to a relatively large sample of Contiguous United States (CONUS) landfalling TCs. A few other studies have examined landfalling TC precipitation patterns and QPF verification in the context of flooding, including Villarini et al. (2011) which examines QPE only, but does include storm relative quadrant analysis. Luitel et al. (2018) examine the large-scale QPF skill of CONUS landfalling TCs and attempt to incorporate observational uncertainty through the use of multiple QPE products. While these studies are disparate in their metrics and methods, one common theme is the lack of advanced spatial verification, such as an integrated leveraging of storm-relative analysis and object-oriented verification tools.

Two recent studies by Chen et al. (2018) and Yu et al. (2020) use object-oriented verification techniques and storm-relative coordinate systems in a coherent framework for a large sample of storms. They use the contiguous rain area methodology (CRA; Ebert and McBride 2000; Ebert and Gallus 2009) to examine location, intensity, and spatial pattern errors in TC QPF. The CRA methodology was one of the first object-oriented rainfall verification methodologies. This methodology uses a best-fit algorithm to shift the forecast region to find the best location match between the forecast and observed precipitation areas. Then the original forecast errors can be decomposed into displacement, pattern, rotation, and volume errors, and aggregated to produce mean error statistics. Given limited TC QPF evaluations, and in particular object-based studies, it is worthwhile for the community to explore additional object-based methods along with tools to perform a range of TC QPF evaluations within one software package.

4

Here we present a set of TC-centric precipitation verification capabilities all of which are implemented in the Model Evaluation Tools (MET; Brown et al. 2021) using verification methods ranging from traditional QPF approaches, to storm-centric coordinates, to feature verification using objects identified with the Method for Object-based Diagnostic Evaluation (MODE; Davis et al. 2006). MODE has been used in previous feature-based verification studies (Gilleland et al. 2009; Clark et al. 2014; Wolff et al. 2014), which provide a baseline level of understanding of its benefits and limitations. MODE is distinctly different from the CRA method, yet complementary. The MODE object identification algorithm was developed to mimic the subjective human forecaster's ability to match observed and forecasted objects and uses a multistep process and a fuzzy logic engine to match and merge objects in the forecast and observation fields (Halley Gotway et al. 2021). MODE considers a wide range of object attributes not considered in the CRA method, and allows for evaluation of both matched and unmatched objects (Davis et al. 2006).

Three different storms were selected for demonstration that exhibit a range of typical North Atlantic basin TCs including landfalling and recurving TCs and both weak and strong TCs. We anonymized the storms avoid focusing on specific storm performance rather than the utility of the methods. The datasets used to demonstrate our TC QPF verification methods are described in Section 2, and the verification methods and tools are described in Section 3. Section 4 steps through example QPF evaluations, and finally we provide some summary thoughts and discussions in Section 5.

2. Data

a. Tropical cyclone track data

The Automated Tropical Cyclone Forecast (ATCF) file format was developed at the Naval Oceanographic and Atmospheric Research Laboratory (NRL) (Miller et al. 1990) and is used by the NHC. This file adheres to an ASCII format that includes common fields that describe TC information such as basin, cyclone number, position, and intensity. ATCF files are generally produced by running a vortex tracking software over the gridded model data to isolate the track

5

location and other relevant storm fields. ATCF file format is also followed for the NHC best track, which is a subjectively smoothed post storm analysis of TC location, maximum intensity and minimum sea-level pressure at 6-hourly intervals, determined retrospectively using all available data (Jarvinen et al. 1984; Rappaport et al. 2009, Landsea et al. 2013). We use the NHC best track dataset as our track location observation to account for position errors in the model fields for these methods.

b. Quantitative Precipitation Estimates (QPE)

The QPE products used for our demonstration include the NOAA Climate Prediction Center (CPC) Morphing technique (CMORPH) QPE product (Joyce et al. 2004) and the National Centers for Environmental Prediction (NCEP) Stage IV QPE product (Baldwin and Mitchell 1997). Due to the radar- and gauge-based nature of Stage IV, CMORPH is used for verification over water, while Stage IV is used for verification over land. In particular, Stage IV is a 4 km multi sensor (ground radar and rain gauge) based QPE product that is a merged analysis produced by the 12 River Forecast Centers (RFCs) individual Stage IV analyses, which includes extensive manual quality control. However, even the Stage IV QPE can be characterized by biases and random error (uncertainty) due to a myriad of factors such as radar beam blockage or miscalibration, incorrect radar Z-R relationship, rain gauge errors, or a lack of nearby radar or rain gauge observations leading to lengthy interpolation distances. However, Stage IV QPE is still one of the better QPE estimates over CONUS and is often considered the reference dataset (e.g., Beck et al. 2019). Stage IV is available in both hourly and 6 hourly accumulations, and we use the 6 hourly accumulations.

The CMORPH QPE is a satellite based QPE product that blends low Earth orbiting passive microwave precipitation estimates and geostationary infrared (IR) cloud top information. The passive microwave precipitation estimates are morphed in time using high spatiotemporal resolution IR information (Joyce et al. 2004). CMORPH data are available at a variety of spatial and temporal resolutions; here we focus on the 3-hourly 0.25 degree data, which are available for all our demonstration storms; in addition, the resolution of these data is closer to the native resolution of the passive microwave QPE of 10-20 km (Joyce et al. 2004). While CMORPH uses relatively high accuracy (over water) passive microwave precipitation retrievals, there are known

6

biases and relatively high uncertainties (relative to *in situ* or well tuned ground radar QPE) in satellite QPE such as CMORPH (AghaKouchak et al. 2011). Additionally, satellite precipitation estimates like CMORPH have a conditional bias, where they often severely underestimate higher precipitation rates (AghaKouchak et al. 2011; Wright et al. 2017).

c. Model-based Quantitative Precipitation Forecasts (QPFs)

We use QPFs from North Atlantic basin forecasts to demonstrate advanced TC QPF verification approaches with the suite of MET tools described above. The model-based forecasts have been anonymized as this paper is meant to present the tools and demonstrate the type of useful information that can be garnered through the application of the tools, thus the specific model is not relevant. Precipitation forecasts of 120 hours (5 days) are taken from 6-hourly model initializations (00, 06, 12, 18 UTC), and QPF accumulations at various forecast lead times (e.g., 12-h, 72-h) are compared to the QPE products.

d. Temporal aggregation of precipitation fields

QPE accumulations are aggregated to appropriate intervals (e.g., 6-h accumulated precipitation) using the MET PCP-Combine tool, while model-based accumulation intervals are computed directly from the model output by differencing the total precipitation accumulation field between specific lead times (e.g. 18-h minus 12-h). Additionally, QPE products and the model QPF are regridded to a common 0.25 degree grid based on the coarsest resolution dataset (CMORPH) using the MET Regrid-Data-Plane tool unless specifically noted before any comparisons are made.

3. Tools

Here we provide an overview of MET and a more detailed description of the specific MET tools applied in this study and their underlying methods. The MET Users Guide contains comprehensive information regarding their use and configuration (Halley Gotway et al. 2021).

a. MET overview

The MET community verification software package (Brown et al. 2021) was developed to serve both the research and operational NWP communities through a state-of-the-art verification

7

software that is modular and adaptable. MET is freely available and supported, providing both traditional and advanced verification metrics, including spatial verification approaches. The structure and modules that make up the components of MET are show in Figure 1. The grey boxes represent file input and output, the dark green ovals show the pre-processing and reformatting tools, the plotting utilities are shown in the light green ovals, the blue ovals are the statistical tools, and the yellow ovals are the aggregations and analysis tools. The components represented in the lower right section of Figure 1 are the TC-specific MET tools. The TC-Dland, TC-Pairs, and TC-Stat tools were the initial TC tools, introduced to replicate the functionality of the NHC verification system. Several additional TC-specific tools have been added to MET in subsequent releases, where the TC-specific tools are defined by the use of ATCF format input files. The TC tools described herein utilize capabilities developed to account for TC track specific considerations while employing other MET capabilities that require input from the gridded forecast fields inputs.
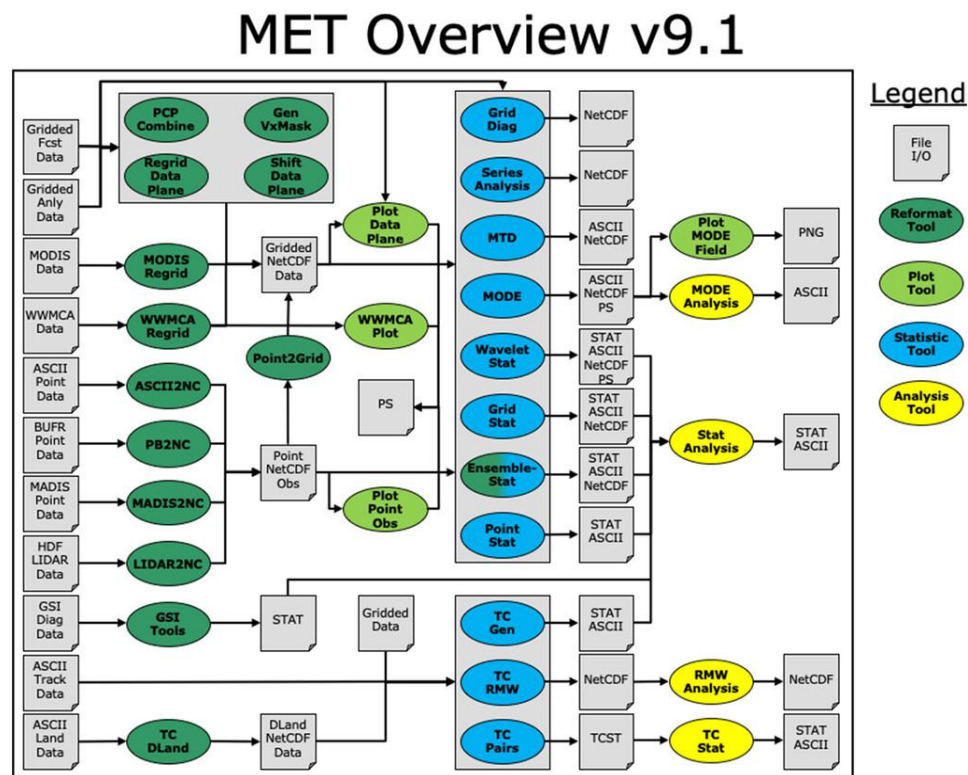

Figure 1. Overview of the structure of the MET package (version 9.1). From Brown et al. (2021), reprinted with permission from American Meteorological Society.

a. Shift data plane (SDP)

8

The SDP tool shifts gridded forecast data to correct for forecast track errors as compared to an observed track. A forecast and observed track of any generic feature of interest could be used, but currently the SDP tool requires the ATCF file format for forecast and observed tracks. Typically, a user would identify a forecast track using some type of vortex tracker (e.g., Marchok 2021), which is used to objectively analyze data to provide an estimate of the storm's central position and track the storm for the duration of the forecast. The corresponding best track analysis (Section 2a), is used for the observations. The SDP tool compares predictions at each individual forecast valid time to the matching observation field and shifts the forecast field using the forecast-observed track difference, essentially following the approach used by Marchok et al. (2007). Figure 2 provides a schematic of a hypothetical forecast shift for three forecast times. Track shifting is useful when track dependent verification is desired through more traditional metrics such as the equitable threat score (ETS) or fractions skill score (FSS; Roberts 2008, Roberts and Lean 2008). Advanced spatial verification techniques such as MODE object-based verification (Davis et al. 2006) account for object displacements, as discussed in Section 3c.
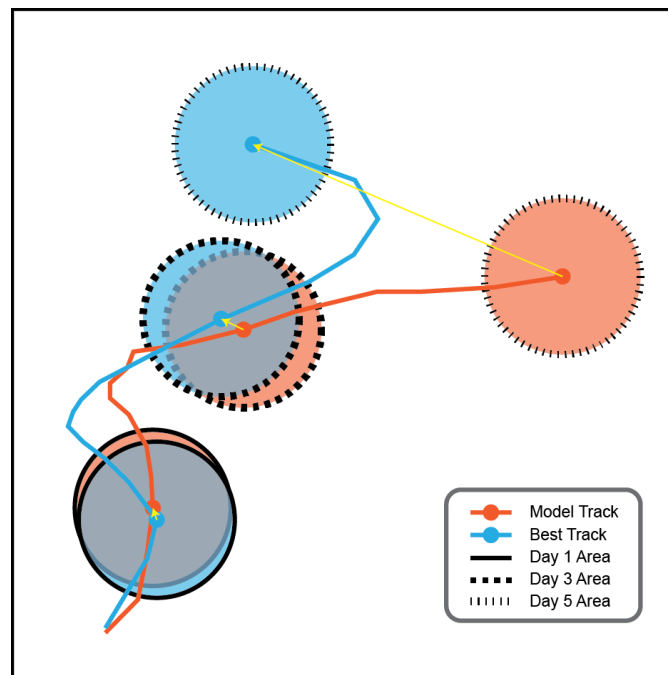

Figure 2. Schematic of forecast field shifting using forecast minus analyzed track errors at three forecast valid times.

b. TC-Radius of Maximum Wind (TC-RMW)

9

MET's TC-radius of maximum wind (TC-RMW) tool regrids TC model data onto a moving range-azimuth grid centered on points along a user-provided storm track, again provided in ATCF format. The radial grid spacing may be set as a factor of the radius of maximum winds (RMW). Figure 3 shows an example storm with the range-azimuth grid in RMW. Transforming forecast and observed data into a common TC-relative coordinate system is useful for storm-relative and process-oriented verification. Many features and processes of TCs are better understood through a physically based storm relative coordinate system (e.g., Marchok et al. 2007; Yu et al. 2015; Cheung et al. 2018).
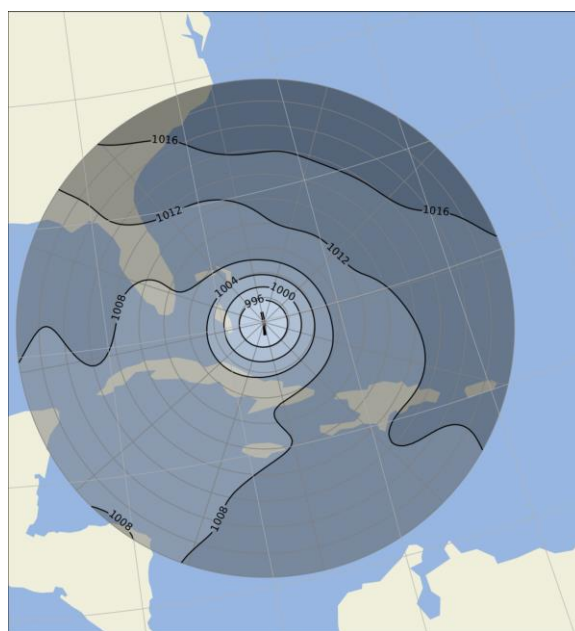

Figure 3. Range-azimuth grid of an example tropical cyclone. The thick black line across the storm center indicates the extent of the radius of maximum winds (RMW) with mean sea level pressure (mb) using filled gray contours for storm context.

c. Method for Object-based Diagnostic Evaluation (MODE)

MODE is an object-oriented spatial verification technique that is designed to recreate subjective human-based object-oriented forecast verification within a quantitative framework. The motivation behind MODE stems from the desire to move beyond traditional forecast verification metrics like probability of detection to assess the spatial information within a forecast in an intuitive manner, and to more appropriately examine errors in gridded forecast values that are impacted by spatial displacements (Davis et al. 2006; Bullock et al. 2016). MODE uses a very simple convolution-threshold process to identify objects where a user-defined circular convolution filter is applied to both gridded forecast and observation fields. Then a user-

10

specified threshold is applied to both convolved fields and objects are identified in the thresholded convolved fields. The user-specified threshold may be different for forecast and observed fields to account for distributional shifts between the two fields. After thresholding, the raw field values are substituted at all grid cells greater than or equal to the threshold value. Object attributes (e.g., centroid location, intensity distribution) are computed from the restored field and can be compared between the matched forecast and observed objects (Davis et al. 2006; Bullock et al. 2016).

A fuzzy logic approach is used to merge objects within a single forecast or observed grid, and to match objects between the forecast and observed grids. MODE uses a combination of object attribute interest values and weights. The object interest functions are defined per object attribute and can be a piecewise linear function or one of several available algebraic expressions in MODE. This function defines what values of an object attribute are considered interesting, and how interesting they are. Attribute weights define which object attributes are most important to the user for the matching process and can be any nonnegative value, as MODE normalizes the user weights during matching. Attributes with large weights are more important in the matching algorithm (Bullock et al. 2016). The combination of interest and weights defines the total interest assigned to a pair of objects, and a user-specified total interest threshold defines what forecast-observation object pairs are considered matches.

Parameter name	Parameter description	Parameter range and description
Convolution Radius	This parameter defines the radius, and thus the size, of the circular convolution filter	A nonnegative number of grid cells
Convolution Threshold	The threshold values applied to the convolved observed and forecast fields that define where the subsequent restored raw fields have valid data	Any valid observation or forecast field value
Interest function[1]	The function that defines object attribute interest magnitude across all possible attribute values	Bounded by [0,1] for all object attribute values
Weight[1]	A scalar value that determines the relative importance of object attributes	Any nonnegative value
Total Interest	Minimum value for forecast-observation object	Bounded [0,1]

		pairs to be considered a match	

Table 1. Description of primary MODE parameters that influence object identification and merging within a forecast or observed field, and matching between forecasts and observations. From Bullock et al. (2016).

[1]The interest functions and weights can be specified for all object attributes computed in MODE.

For each object that MODE identifies in either the forecast or observed field, characteristics of the objects are computed including their area, centroid locations, intensity statistics, which in this case are based on the grid cell precipitation accumulations within the object, and the complexity of the object outline contour. For forecast-observation matched objects, ratios of these quantities are computed, as well as contingency table statistics based on the objects (Halley Gotway et al. 2021). After MODE output is generated, the MODE-Analysis or other user developed processing tools can be used to aggregate the aforementioned statistics. The ratios of matched forecast-observation object pairs measure multiplicative biases of those quantities when aggregated. We use a combination of additional processing of the MODE matched pairs and the MODE-Analysis tool to identify matched objects and compute forecast/observation ratios such that values larger than 1 indicate an over forecast and values smaller than 1 indicate an under forecast of a given metric.

4. Application Examples

This section steps through several applications of the aforementioned tools. Workflow diagrams for the application examples are provided in the Appendix.

a. Storm-relative verification

Storm relative QPF verification is useful for improving both model process representation and understanding model biases for estimating rainfall flooding due to landfalling TCs (Marchok et al. 2007; Cheung et al. 2018; Yu et al. 2020). Storm-relative verification can be performed within MET via a combination of regional masking capabilities using the Gen-Vx-Mask tool and any of the MET analysis tools such as the Grid-Stat or Series-Analysis tools. Furthermore, both unshifted and shifted gridded output using the Shift-Data-Plane tool can be ingested into any MET analysis tool. An example workflow diagram is provided in Figure A1. Figure 4 presents a schematic of storm-relative distance masks for both model and observations that can be used to

12

generate storm-relative verification statistics. The model (Fig 4a) is a 168-h forecast track for a single initialization and the corresponding observations (Fig 4b) for each forecast valid time.
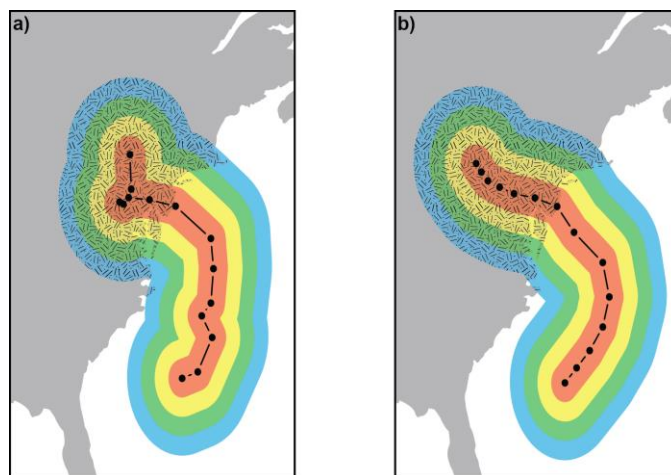

Figure 4. Schematic of storm-relative distance masks within MET. User-specified range intervals (100 km) shown in colors are computed relative to the storm center (black line with circle markers, 12-h interval between markers) using the Gen-Vx-Mask Tool for both the a) model and b) observations. Additional masking between the land (hatching colors over gray background) and water (colors over white background) is highlighted here.

An example of total forecast accumulated precipitation at each gridpoint for one forecast initialization time of storm A is highlighted in Figure 5a. The combination of multiple observational products is also demonstrated through the use of CMORPH precipitation over oceanic pixels (Fig. 5b) and Stage IV precipitation over land pixels (Fig 5c). Distributions of total precipitation accumulations across forecast lead times relative to the forecast and observed tracks aggregated across model initialization times highlight overall model precipitation placement tendencies as compared to observations. Boxplots are generated for all precipitation datasets for each distance band using all grid points in each band (Fig 5d). Here the forecast has a tendency to underestimate the amount of precipitation over water, particularly within 200 km of the storm center as compared to CMORPH. However, when the storm is over land, the model generates much higher precipitation rates and larger total area of precipitation (roughly 7% larger, 20,597 versus 19,121 points, Fig. 5d) within 200 km of the storm center. Similar to Figure 5, we aggregate all matched forecast and observation distributions of accumulated precipitation for each forecast initialization for storm A to generate Figure 6. For 37 forecast initializations, the same forecast trends are seen (Fig. 6). There is a large over estimation of precipitation accumulation and precipitation area (roughly 30% larger area in the model, 490,240 versus

13

383,063 points) within 200 km of the storm center over land, while over water the accumulation distributions and total areas are very similar (712,751 versus 787,839 points). Other traditional metrics such as contingency tables and skill scores can be computed using MET on the shifted grid files, and are extensively documented and discussed in the MET and METplus Users' Guides (Halley Gotway et al. 2021; Win-Gildenmeister et al. 2021) and literature (e.g., Wilks 2019).
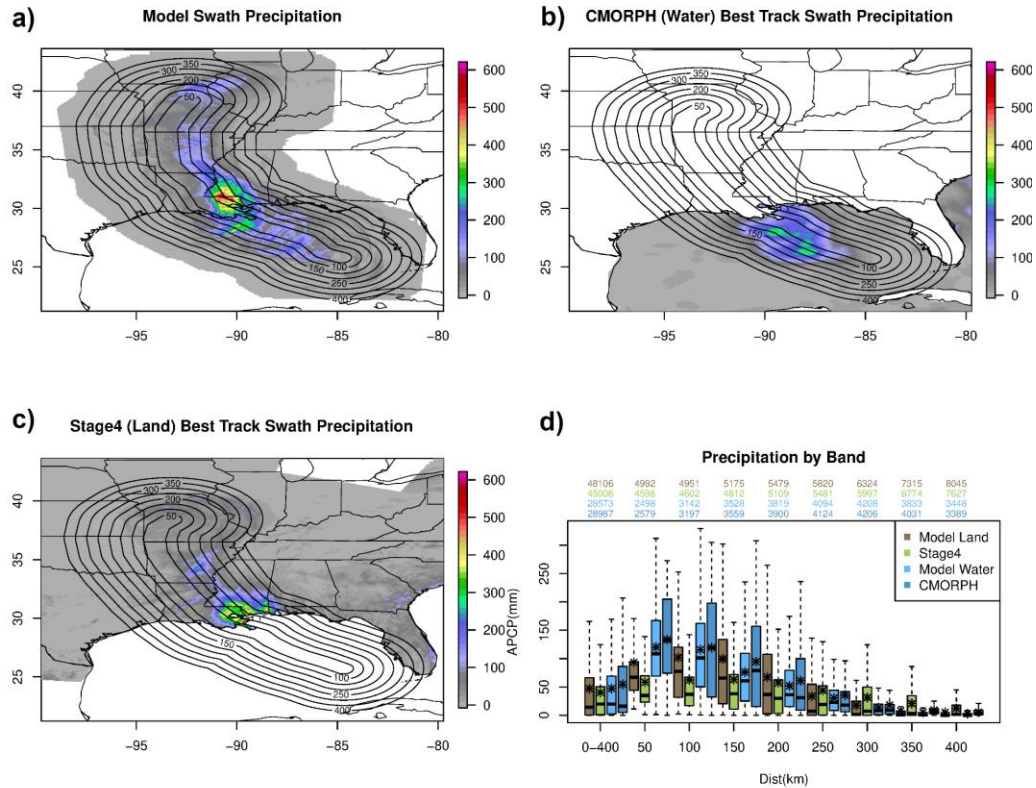

Figure 5. Example total forecast precipitation accumulation (mm) in color shading and distance from storm center at 50 km intervals for a single initialization of storm A for a) model, b) observed best track over water with CMORPH, c) observed best track over land with Stage IV, and d) summary boxplot of precipitation accumulation (mm). For each box, the bold line indicates the median, the mean of the distribution is depicted as an asterisk, and the bottom and top edges show the 25th and 75th percentiles, respectively. The whiskers extend to the most extreme data points not considered outliers. The number of non-zero precipitation grid cells is listed above each range bin. This is a model forecast including land and water with the corresponding CMORPH observations being masked to only include water pixels and Stage IV observations only land pixels. Note the first column in d) is for the full 0-400 km distance.
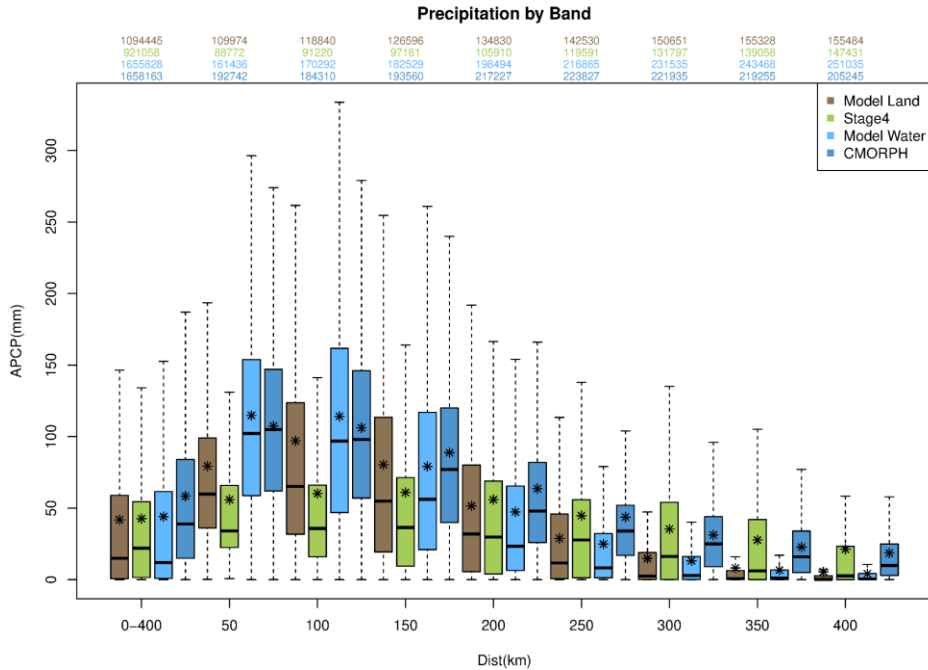
14

Figure 6. Boxplot of total forecast precipitation accumulation (mm) aggregated across 37 forecast initialization times from storm A for a model over land and water and corresponding observations with CMORPH being masked to only include water pixels and Stage IV only land pixels. Note the first column is for the full 0-400 km distance. For each box, the bold line indicates the median, the mean of the distribution is depicted as an asterisk, and the bottom and top edges show the 25th and 75th percentiles, respectively. The whiskers extend to the most extreme data points not considered outliers.

b. RMW-based verification

Moving beyond storm-relative verification on traditional cartesian coordinates into storm-relative cylindrical coordinates scaled by the RMW can provide additional insights into model biases and process deficiencies. The MET TC-RMW tool can be used for these evaluations, as shown in Figure A2. Figure 7a shows a composite of 6-hourly precipitation accumulation for 34 12-h forecasts for storm B, with the corresponding CMORPH observations shown in Fig. 7b. This approach to displaying the data highlights the differences between the forecasted and observed storm structure. The modeled storm in this case is more compact than the observed one, and has a larger precipitation gradient away from the storm center, with the most intense precipitation occurring within about two times the RMW. It is also more symmetric than observed with larger precipitation accumulations in the southern semicircle within five times the RMW. The observed storm has a smaller precipitation gradient with larger accumulations

15

outside two times the RMW in the northeast quadrant, and a lack of precipitation in the southern semicircle. There is also a notable time-mean precipitation maxima extending radially away from the storm center just to the north of due East (90 degrees) in the observed storm that is absent in the modeled storm. Given the known underestimation bias associated with CMORPH for higher precipitation rates, the differences between forecast and actual precipitation rates near the center of the storm may not be as drastic as shown in Fig 7, but the general spatial structure comparison should not be impacted by this bias.
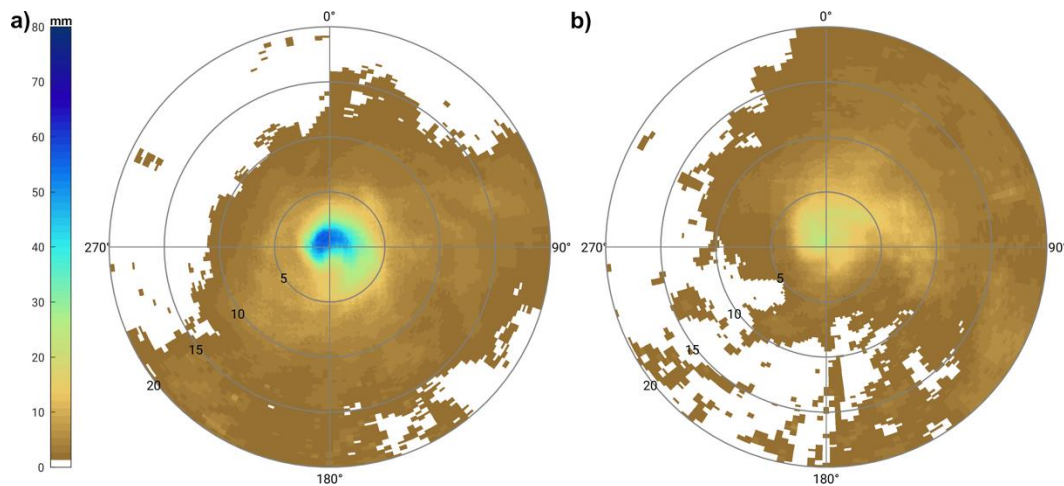
Figure 7. Mean six-hourly precipitation accumulation (mm) for 34 initializations of storm B at the 12-hour lead time for a) model and b) observations at the same valid times. Radii are proportional to the radius of maximum wind.

The TC-RMW output can also be aggregated across time and space into distributions relative to the storm center similar to those shown in Section 4a. Because TC-RMW is currently lacking land-ocean masking capabilities, storm C, which does not make landfall, is used in Figure 8. However, similar trends are observed given the output is from the sample model and compared to the same observation type (CMORPH). The model consistently produces larger-than-observed precipitation totals irrespective of distance from the storm center, with a more pronounced bias near the center, which may be an artifact of CMORPH's known underestimation bias for heavy precipitation. Care must be taken when aggregating across multiple initializations to take into account for factors like TC developmental stage, based on the user's evaluation goals.
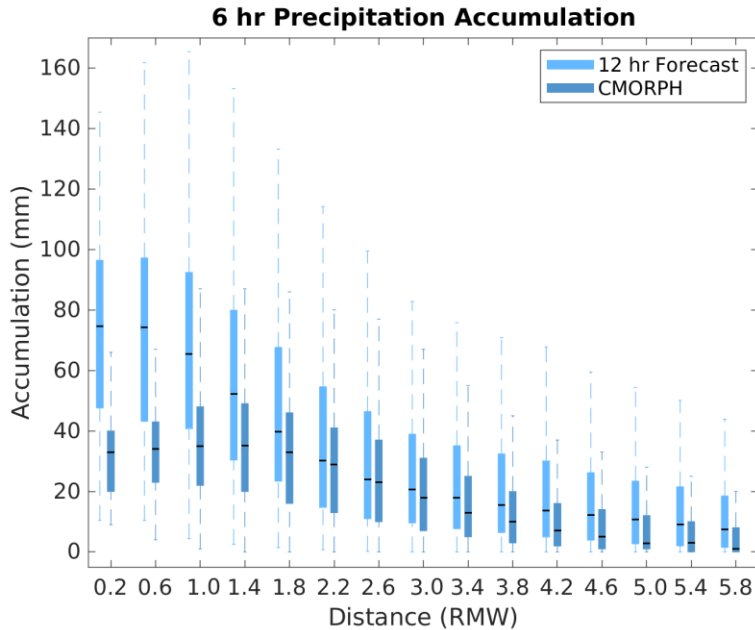
16

Figure 8. Boxplot of six-hourly precipitation accumulation (mm) aggregated across 47 forecast initialization times from storm C for the 12-hour lead time forecasts. For each box, the bold line indicates the median and the bottom and top edges show the 25th and 75th percentiles, respectively. The whiskers extend to the most extreme data points not considered outliers.

c. Object-based verification

As discussed in Section 3, the MODE object-based verification approach can be used to identify individual objects or object groups in both the forecast and observation fields, and then identify matched objects or object groups that can be compared between the forecast and observations. A workflow diagram demonstrating the example below for running MODE is provided in Figure A3. An example of a 6-hourly accumulated precipitation forecast object group and observed precipitation object group for landfalling storm A is shown in Figure 9. In this case, we use masking capabilities within MET to focus on areas with valid Stage IV data. When using masks, only pixels within the valid masking region will be considered, thus the object will be split or not matched if the forecast or observation object falls entirely outside the mask. Note that the size and location of the objects depends on the threshold used to define the objects. The Shift-Data-Plane tool was applied to the forecast data to correct for TC track errors. Shifting the tracks before running MODE may be useful to assure that model and observed TC features are properly matched, rather than TC features matching to non-TC features in the cases with large track errors. We see a group of forecasted objects highlighting the rain shield of a weakening TC, with the primary object being larger than the observed rain shield object.

17

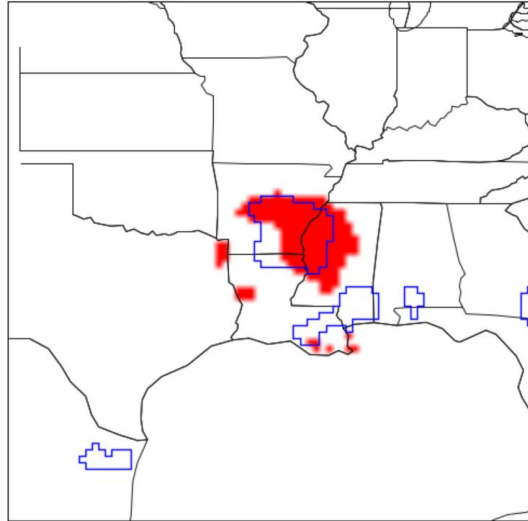**Forecast Objects with Observation Outlines**

Figure 9. An example of MODE accumulated precipitation forecast objects for storm A, identified as one object cluster in red with observation objects overlaid using blue outlines.

Area, complexity, and intensity ratio distributions for 6-hourly precipitation forecasts at 12- and 72-hour forecast lead times are shown in Figure 10. Here we see that the model has a tendency to have larger-than-observed objects both over land and water, higher complexity over land, with larger biases over land than over water. The intensity bias increases over land with lead time in the two storms that make landfall. The combination of larger and more intense objects implies that this model overpredicts total rainfall over land, and possibly would overpredict inland freshwater flooding impacts, although inland flooding is controlled by additional processes within the landscape.
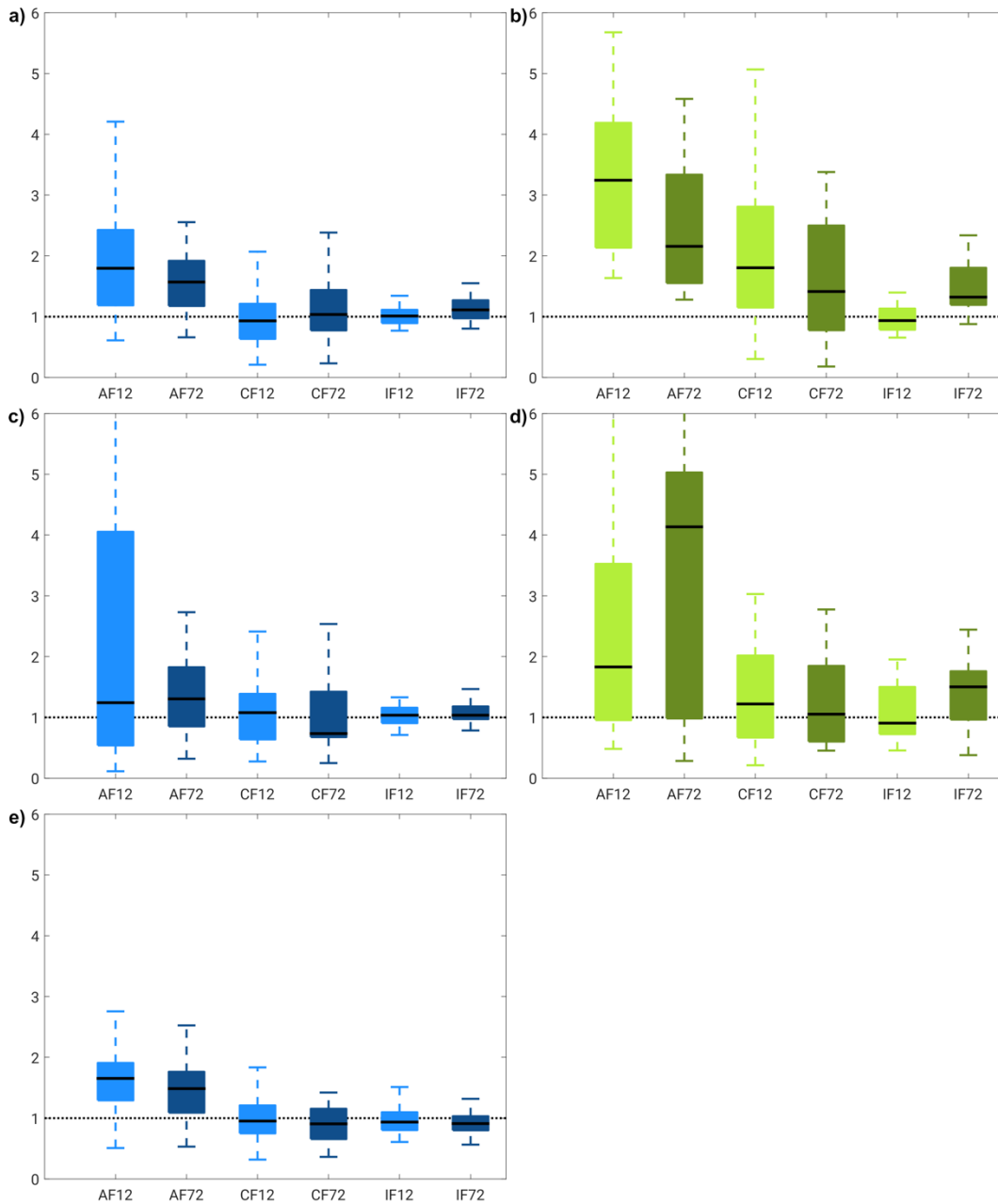
Figure 10. Aggregated MODE object statistics for all 6-hourly precipitation accumulations at 12- and 72-hour forecast lead times for the three example storms segregated by land (b,d) and water (a,c,e) masks. a-b) show storm A, c-d) show storm B, and e) shows storm C which never interacted with land. The x-axis abbreviations are defined as the first letter indicates the object statistic, A=area ratio, C=complexity ratio, and I=Intensity ratio, while F12 and F72 indicates the forecast lead time in hours. All ratios are forecast/observation.

Finally, object composites (using the MODE computed object centroids as the reference point) are useful to identify spatial patterns in model biases. Figure 11 displays the composite object for the observation, model, and difference (model-observations) field for 6-hour accumulation at 12-

hour forecast lead times from storm C. The model in this case has a much larger storm than was observed, roughly 1.75 times larger on average (Fig. 10e), and this is clearly seen in the composite object when comparing Figure 11a to 11b. The model also produces more total rainfall associated with the larger object, and also higher mean accumulation within the storm object (Fig. 11c). One caveat is that the CMORPH observations almost certainly underestimate precipitation accumulations in intense convection. Note that storm C only occurs over water, and therefore only CMORPH observation objects are included in the composite for Figure 11b. In the case of landfalling storms, a combined observation grid (e.g. CMORPH over water and Stage IV over land) would be necessary for identifying observation objects.
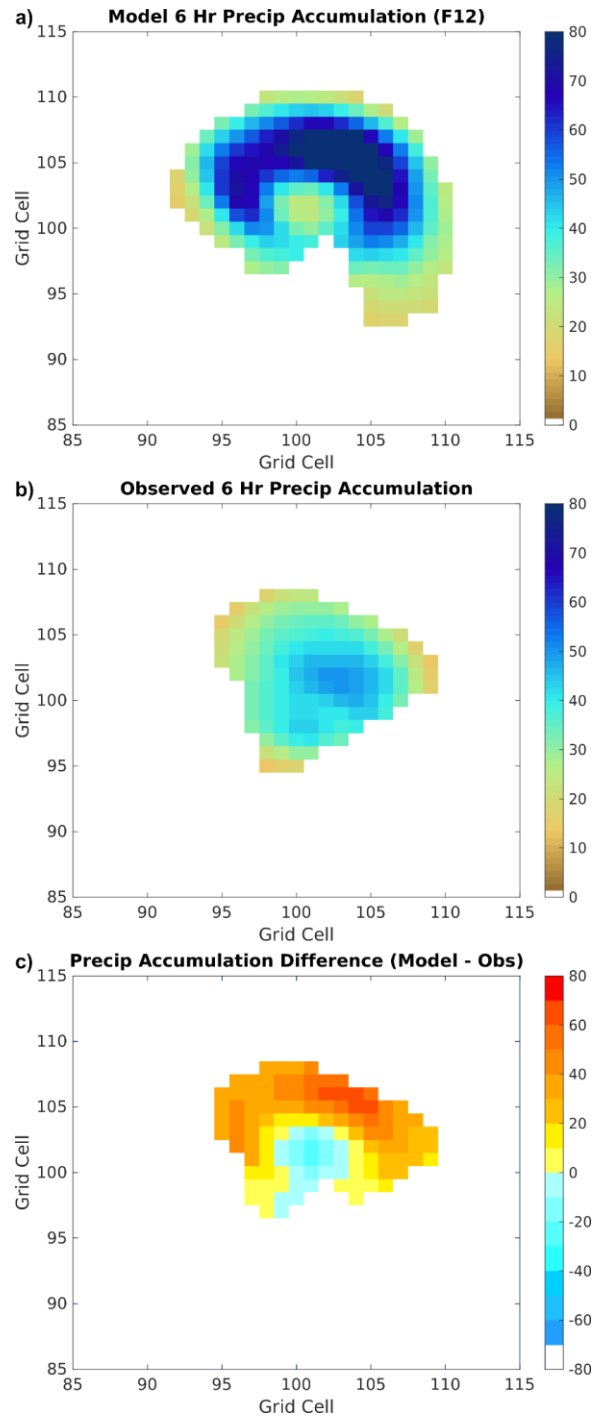
20

Figure 11. Composite of 6-hour accumulated precipitation (mm) MODE objects of storm C across 20 12-hour forecasts for a) model forecasts, b) CMORPH observations, and c) the model-observations difference field.

5. Summary and Discussion

Moving beyond traditional point- or grid-based model QPF verification is useful for improving understanding of model biases that affect TC impacts such as inland freshwater flooding and also for building capabilities for process-based model diagnostics (Marchok et al. 2007; Villarini et al. 2011; Cheung et al. 2018; Chen et al. 2018). However, relatively few studies exist that have applied diagnostic and advanced verification methods to TC QPF verification. The MET software now contains a set of generalized and TC-specific spatial verification tools that allow for meaningful TC QPF verification along the TC tracks. Storm-relative verification using observed track-shifted (or unshifted) model data in absolute distance (km) can be useful for identifying model biases in precipitation accumulation in relation to TC centers, which have implications, for example, for the interpretation and application of forecasts of landfalling TCs (Figs. 4-6). Moving towards RMW-based verification using TC-RMW adds additional storm relative capabilities of regridding QPF and QPE into polar coordinates (Figs. 7-8). This approach is a process-based model diagnostic and verification technique that is a powerful tool to reveal feedbacks between forecast tracks, intensity and precipitation distributions (e.g. Yu et al. 2015; Cheung et al. 2018). Object-based verification including complex land masking capabilities provides even more nuanced verification capabilities. Precipitation objects of interest, either the central core of TCs, or extended areas of rainfall after landfall, can be identified, matched to observations and quickly aggregated to build meaningful spatial and summary verification statistics (Figs. 9-11). Finally, future work could explore the detailed differences between traditional and advanced spatial methods specific to TCs similar to what has been done for other precipitation features (Davis et al. 2006; Wolff et al. 2014).

Within this analysis, a few points deserve more detailed attention. First, gridded observations of precipitation are inherently highly uncertain and often contain complex bias structures as noted here. Even *in situ* precipitation observations have uncertainties and issues with point-to-grid representativeness. Thus, future work should include examination of multiple observational products (e.g. IMERG/Global Satellite Mapping of Precipitation (GSMaP), Multi-Radar/Multi-Sensor System (MRMS), Climatology-Calibrated Precipitation Analysis (CCPA)) in an effort to understand and potentially constrain observation uncertainty and interpret model performance statistics. Second, MODE is highly configurable with several tunable parameters that control object identification and matching (Table 1). In this work we tested many different convolution

thresholds and convolution radii and chose an optimal combination for this specific model and observation set that would identify similar objects in the model output and both CMORPH and Stage IV observations through visual inspection of MODE summary graphics. For any given model and observation dataset, and application, the optimal convolution threshold and radii will likely vary. Additionally, the matching scheme options may also be modified to influence how objects are grouped within model or observation fields, and how they are matched between the model and observations. Therefore, we strongly recommend testing a variety of MODE configurations and visually examining the output for expected behavior for a specific application. This requires a strong understanding of the types of verification questions the user is trying to answer before undertaking an in-depth verification effort. Finally, an expansion of the ability to use RMW units within other MET tools, such as MODE and other gridded data based analysis tools, will be considered for future MET releases to enhance the process-diagnostic utility of the toolkit.

Data Availability Statement

Due to privacy and ethical concerns, neither the data nor the source of the data can be made freely available. Please contact the corresponding author to discuss access to the data.

APPENDIX

Workflow Diagrams

Here we present workflow diagrams for the three general examples of TC QPF verification using MET tools to aid interested users. Figures A1-A3 correspond to the discussion in sections 4a-c, respectively.
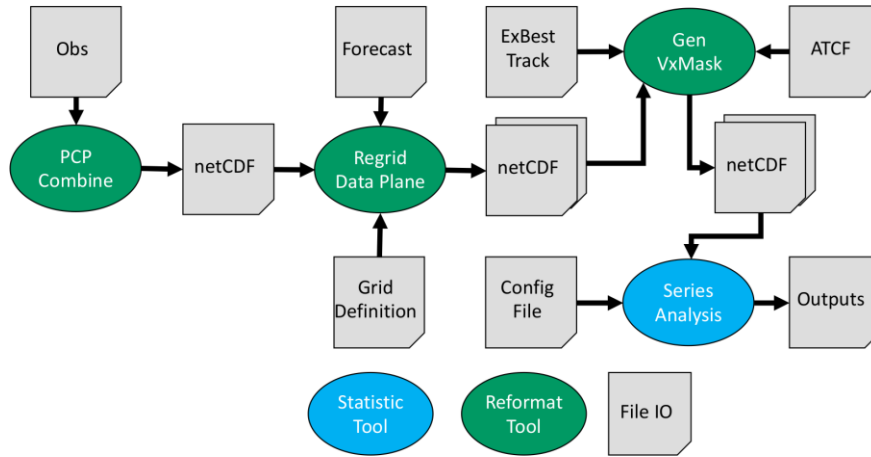

Figure A1. Workflow diagram example for creating a storm-relative distance verification using Series Analysis. Shift Data Plane needs to be run for the observations and forecast individually. The diagram components follow Figure 2 in Brown et al. (2021).
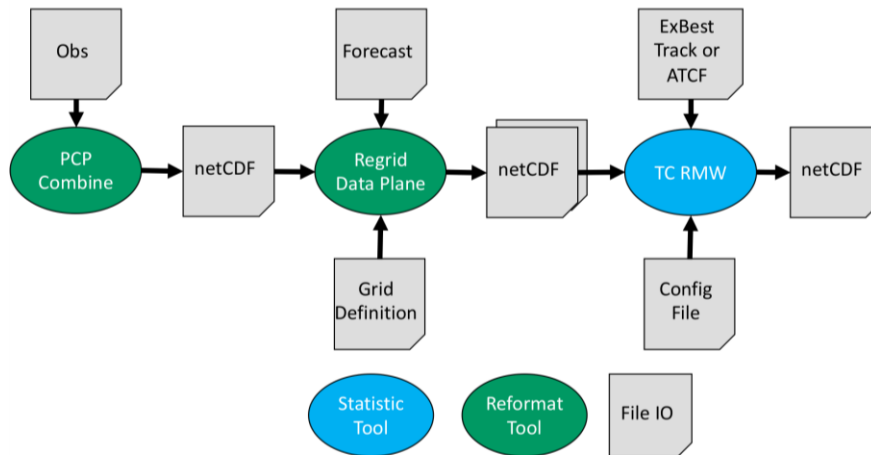

Figure A2. Workflow diagram example for creating a TC-RMW output file for model and observed fields. Note this requires TC-RMW to be run twice, once for the forecast and once for the observations. Diagram components follow Figure 2 in Brown et al. (2021).
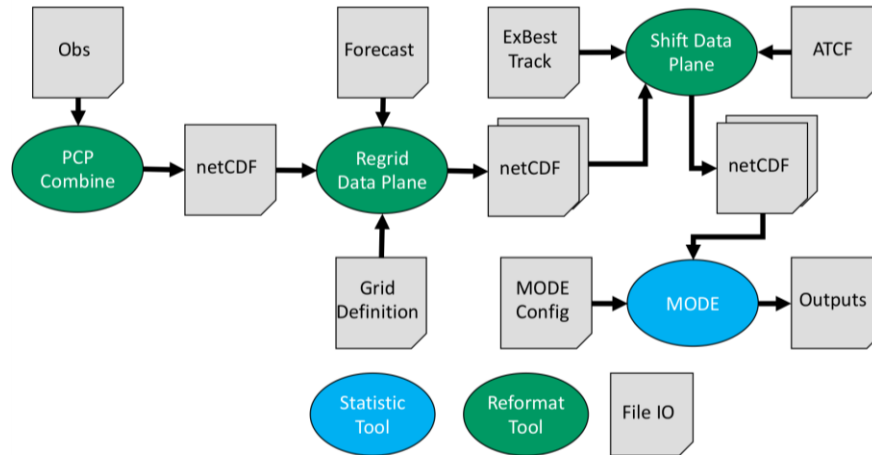
24

Figure A3. Workflow diagram for example for performing MODE analysis. Shift Data Plane needs to be run for the observations and forecast individually. Diagram components follow Figure 2 in Brown et al. (2021).

REFERENCES

AghaKouchak, A., A. Behrangi, S. Sorooshian, K. Hsu, and E. Amitai, 2011: Evaluation of satellite-retrieved extreme precipitation rates across the central United States. *J. Geophys. Res.*, **116**, D02115, doi:10:1029/2010JD014741.

Baldwin, M. E., and K. E. Mitchell, 1997: The NCEP hourly multisensor U.S. precipitation analysis for operations and GCIP research. Preprints, *13th Conf. on Hydrology,* Long Beach, CA, Amer. Meteor. Soc., 54–55.

Beck, H. E., M. Pan, T. Roy, G. P. Weedon, F. Pappenberger, A. I. J. M. van Dijk, G. J. Huffman, R. F. Adler, and E. F. Wood, 2019: Daily evaluation of 26 precipitation datasets using Stage-IV gauge-radar data for the CONUS, *Hydrol. Earth Syst. Sci.*, **23**, 207–224, https://doi.org/10.5194/hess-23-207-2019.

Biswas, M., and Coauthors, 2018: Hurricane Weather Research and Forecasting (HWRF) Model: 2017 scientific documentation. NCAR Tech. Note NCAR/TN-544+STR, 111 pp., https://doi.org/10.5065/D6MK6BPR.

Brown, B., T. Jensen, J. H. Gotway, R. Bullock, E. Gilleland, T. Fowler, K. Newman, D. Adriaansen, L. Blank, T. Burek, M. Harrold, T. Hertneky, C. Kalb, P. Kucera, L. Nance, J. Opatz, J. Vigh, and J. Wolff, 2021: The Model Evaluation Tools (MET): More than a Decade of

Community-Supported Forecast Verification, *Bulletin of the American Meteorological Society*, **102**(4), E782-E807.

Bullock, R. G., B. G. Brown, and T. L. Fowler, 2016: Method for Object-Based Diagnostic Evaluation. NCAR Tech. Note NCAR/TN-532+STR, doi:10.5065/D61V5CBS

Chen, Y., E. E. Ebert, N. E. Davidson, and K. J. E. Walsh, 2018: Application of contiguous rain area (CRA) methods to tropical cyclone rainfall forecast verification. *Earth and Space Science*, **5**, 736–752. https://doi.org/10.1029/2018EA000412

Cheung, K., Z. Yu, R. L. Elsberry, M. Bell, H. Jiang, T. C. Lee, K. C. Lu, Y. Oikawa, L. Qi, R. F. Rogers, and K. Tsuboki, 2018: Recent advances in research and forecasting of tropical cyclone rainfall. *Tropical Cyclone Research and Review*, **7**(2), 106-127

Clark, A. J., R. G. Bullock, T. L. Jensen, M. Xue, and F. Kong, 2014: Application of object–based time–domain diagnostics for tracking precipitation systems in convection–allowing models. *Weather and Forecasting*, **29**(3), 517–542. https://doi.org/10.1175/WAF-D-13-00098.1

Clark, P., N. Roberts, H. Lean, S. P. Ballard, and C. Charlton- Perez, 2016: Convection-permitting models: a step- change in rainfall forecasting. *Meteorological Applications*, **23**(2), pp.165-181.

Davis, C. A., B. G. Brown, and R. G. Bullock, 2006: Object-based verification of precipitation forecasts. Part I: Methodology and application to mesoscale rain areas. *Mon. Wea. Rev.*, **134**, 1772–1784, https://doi.org/10.1175/MWR3145.1.

DeMaria, M., J. A. Knaff, and B. H. Connell, 2001: A Tropical Cyclone Genesis Parameter for the Tropical Atlantic, *Weather and Forecasting*, **16**(2), 219-233.

Ebert, E. E., and W. A. Gallus, 2009: Toward better understanding of the contiguous rain area (CRA) method for spatial forecast verification. *Weather and Forecasting*, **24**(5), 1401–1415. https://doi.org/10.1175/2009WAF2222252.1

Ebert, E., and J. McBride, 2000: Verification of precipitation in weather systems: Determination of systematic errors. *Journal of Hydrology*, **239**(1–4), 179–202. https://doi.org/10.1016/S0022-1694(00)00343-7

ECMWF, 2022: Implementation of IFS Cycle 48r1. ECMWF, https://confluence.ecmwf.int/display/FCST/Implementation+of+IFS+Cycle+48r1, last accessed on 8 March 2022.

Franklin, J. L., C. J. McAdie, and M. B. Lawrence, 2003: Trends in Track Forecasting for Tropical Cyclones Threatening the United States, 1970–2001, *Bulletin of the American Meteorological Society*, **84**(9), 1197-1204.

Gall, R., J. Franklin, F. Marks, E. N. Rappaport, and F. Toepfer, 2013: The Hurricane Forecast Improvement Project, *Bulletin of the American Meteorological Society*, **94**(3), 329-343.

Gilleland, E., D. Ahijevych, B. Brown, B. Casati, and E. Ebert, 2009: Intercomparison of spatial forecast verification methods. *Weather and Forecasting*, **24**, 1416–1430. https://doi.org/10.1175/2009WAF2222269.1

Goerss, J. S., 2000: Tropical cyclone track forecasts using an ensemble of dynamical models. *Mon. Wea. Rev.*, **128**, 1187–1193.

Haddad, Z. S., E. A. Smith, C. D. Kummerow, T. Iguchi, M. R., Farrar, S. L. Durden, M. Alves, and W. S. Olson, 1997: The TRMM "day-1" radar/radiometer combined rain-profiling algorithm. *J. Meteor. Soc. Japan*, **75**, 799–809.

Haiden, T., B. Casati, C. Cohelo, E. Gilleland, R. Ashrit, M. Dorninger, and C. Marsigli, 2019: Process-oriented verification. *WMO WWRP Joint Working Group for Forecast Verification Research.* http://wgne.meteoinfo.ru/wp-content/uploads/2019/05/ProcessVerif_JWGFVR_final.pdf Retrieved on 8 March 2022

Halley Gotway, J., K. Newman, H. Soh, J. Opatz, T. Jensen, J. Prestopnik, L. Goodrich, D. Fillmore, B. Brown, R. Bullock, T. Fowler, 2021: The MET Version 10.0.1 User's Guide. Developmental Testbed Center. Available at: https://github.com/dtcenter/MET/releases

Halperin, D. J., H. E. Fuelberg, R. E. Hart, J. H. Cossuth, P. Sura, and R. J. Pasch, 2013: An evaluation of tropical cyclone genesis forecasts from global numerical models. *Wea. Forecasting*, **28**, 1423–1445, doi:10.1175/WAF-D-13-00008.1.

Halperin, D. J., R. E. Hart, H. E. Fuelberg, and J. H. Cossuth, 2017: The Development and Evaluation of a Statistical–Dynamical Tropical Cyclone Genesis Guidance Tool, *Weather and Forecasting*, **32**(1), 27-46.

Hemming, J. T., 2017: Tropical cyclone tracking and verification techniques for Met Office numerical weather prediction models. *Meteorological Applications*, **24**, 1-8, doi:10.1002/met.1599.

Huffman, G. J., 2006: Satellite-based estimation of precipitation using microwave sensors. Encyclopedia of Hydrological Sciences, M. G. Anderson, Ed., John Wiley and Sons, 965–980.

Huffman, G. J., D. T. Bolvin, D. Braithwaite, K.-L. Hsu, R. J. Joyce, C. Kidd, E. J. Nelkin, S. Sorooshian, E. F. Stocker, J. Tan, D. B. Wolff, and P. Xie, 2020: Integrated Multi-satellite Retrievals for the Global Precipitation Measurement (GPM) Mission (IMERG). *Satellite Precipitation Measurement. Advances in Global Change Research,* vol 67. Springer, Cham. https://doi.org/10.1007/978-3-030-24568-9_19

Jarvinen, B. R., C. J. Neumann, and M. A. S. Davis, 1984: A tropical cyclone data tape for the North Atlantic Basin, 1886–1983: Contents, limitations, and uses. NOAA Tech. Memo. NWS NHC 22, Coral Gables, FL, 24 pp.

Joyce, R. J., J. E. Janowiak, P. A. Arkin, and P. Xie, 2004: CMORPH: A method that produces global precipitation estimates from passive microwave and infrared data at high spatial and temporal resolution. *J. Hydromet.*, **5**, 487-503.

Kim, D., Y. Moon, S. J. Camargo, A. A. Wing, A. H. Sobel, H. Murakami, G. A. Vecchi, M. Zhao, and E. Page, 2018: Process-oriented diagnosis of tropical cyclones in high-resolution GCMs. *J. Climate,* **31**, 1685-1702.

Landsea, C. W., and J. L. Franklin, 2013: Atlantic Hurricane Database Uncertainty and Presentation of a New Database Format. *Mon. Wea. Rev.,* **141,** 3576-3592, https://doi.org/10.1175/MWR-D-12-00254.1.

Lin, Y., and K. E. Mitchell, 2005: The NCEP stage II/IV hourly precipitation analyses: Development and applications. Preprints, *19th Conf. on Hydrology*, San Diego, CA, Amer. Meteor. Soc., 1.2. Available online at https://ams.confex.com/ams/pdfpapers/83847.pdf.

Luitel, B., G. Villarini, and G. A. Vecchi, 2018: Verification of the skill of numerical weather prediction models in forecasting rainfall from US landfalling tropical cyclones. *Journal of hydrology*, **556**, 1026-1037.

Marchok, T., 2021: Important Factors in the Tracking of Tropical Cyclones in Operational Models, *Journal of Applied Meteorology and Climatology*, **60**(9), 1265-1284.

Marchok, T., R. Rogers, and R. Tuleya, 2007: Validation schemes for tropical cyclone quantitative precipitation forecasts: Evaluation of operational models for US landfalling cases. *Weather and forecasting*, **22**(4), 726-746.

Meléndez- Landaverde, E. R., M. Werner, and J. Verkade, 2020: Exploring protective decision-making in the context of impact- based flood warnings. *Journal of Flood Risk Management*, **13**(1), e12587.

Miller, R. J., A. J. Schrader, C. R. Sampson, and T. L. Tsui, 1990: The Automated Tropical Cyclone Forecasting System (ATCF). Wea. Forecasting, **5**, 653-660.

Morrow, B. H., and J. K. Lazo, 2015: Effective tropical cyclone forecast and warning communication: Recent social science contributions. *Tropical Cyclone Research and Review*, **4**(1), 38-48.

National Hurricane Center, 2022: National Hurricane Center Verification. https://www.nhc.noaa.gov/verification/ Retrieved on 8 March 2022.

Qi, W., B. Yong, J. Gourley, 2021: Monitoring the super typhoon lekima by GPM-based near-real-time satellite precipitation estimates. *Journal of Hydrology,* **603**, 126968

Rappaport, E. N., 2014: Fatalities in the United States from Atlantic Tropical Cyclones: New Data and Interpretation, *Bulletin of the American Meteorological Society*, **95**(3), 341-346.

Rappaport, E. N., and co-authors, 2009: Advances and Challenges at the National Hurricane Center, *Weather and Forecasting*, **24**(2), 395-419.

Roberts, N., 2008: Assessing the spatial and temporal variation in the skill of precipitation forecasts from an NWP model. *Meteor. Appl.*, **15 ,** 163–169.

Roberts, N. M., and H. W. Lean, 2008: Scale-Selective Verification of Rainfall Accumulations from High-Resolution Forecasts of Convective Events, *Monthly Weather Review*, *136*(1), 78-97.

Sampson, C. R., J. L. Franklin, J. A. Knaff, and M. DeMaria, 2008: Experiments with a simple tropical cyclone intensity consensus. *Wea. Forecasting*, **23**, 304–312.

Villarini, G., J. A. Smith, M. L. Baeck, T. Marchok, and G. A. Vecchi, 2011: Characterization of rainfall distribution and flooding associated with US landfalling tropical cyclones: analyses of Hurricanes Frances, Ivan, and Jeanne (2004). *J. Geophys. Res.: Atmosph.* 1984–2012, **116**.

Wilks, D., 2019: Statistical Methods in the atmospheric sciences. 4th edition, ISBN-13: 978-0128158234, Elsevier, Cambridge, MA, USA.

Win-Gildenmeister, M., G. McCabe, J. Prestopnik, J. Opatz, J. Halley Gotway, T. Jensen, J. Vigh, M. Row, C. Kalb, H. Fisher, L. Goodrich, D. Adriaansen, J. Frimel, L. Blank, T. Arbetter, 2021: The METplus Version 4.0.0 User's Guide. Developmental Testbed Center. Available at: https://github.com/dtcenter/METplus/releases.

Wright, D. B., D. B. Kirschbaum, and S. Yatheendradas, 2017: Satellite precipitation characterization, error modeling, and error correction using censored shifted gamma distributions. *J. Hydrometeorology*, **18**, 2801-2815.

Wolff, J. K., M. Harrold, T. Fowler, J. H. Gotway, L. Nance, and B. G. Brown, 2014: Beyond the Basics: Evaluating Model-Based Precipitation Forecasts Using Traditional, Spatial, and Object-Based Methods, *Weather and Forecasting*, **29**(6), 1451-1472.

Yang, F., and co-authors, 2020: GFSV16: Further advancements to the UFS medium range weather application in 2021. Unified Forecast System (UFS) Users' workshop, 27-29 July 2020.

https://dtcenter.org/sites/default/files/events/2020/1-fanglin-yang.pdf, last accessed 8 March 2022.

Yu, Z., Y. Wang and H. Xu, 2015: Observed rainfall asymmetry in tropical cyclones making landfall over China. *J. Appl. Meteor. Climatol.*, **54**, 117–136.

Yu, Z., Y. J. Chen, B. Ebert, N. E. Davidson, Y. Xiao, H. Yu, and Y. Duan, 2020: Benchmark rainfall verification of landfall tropical cyclone forecasts by operational ACCESS-TC over China. *Meteorological Applications*, **27**(1), e1842.