JAMES | Journal of Advances in Modeling Earth Systems

RESEARCH ARTICLE
10.1029/2023MS004008

A Potential Vorticity Diagnosis of Tropical Cyclone Track Forecast Errors

Tyler W. Barbero[1], Michael M. Bell[1], Jan-Huey Chen[2], and Philip J. Klotzbach[1]

[1]Department of Atmospheric Science, Colorado State University, Fort Collins, CO, USA, [2]National Oceanic and Atmospheric Administration/Geophysical Fluid Dynamics Laboratory, Princeton, NJ, USA

Key Points:

- Contributions to tropical cyclone movement from individual synoptic systems are quantified using piecewise potential vorticity inversion
- Forecast errors of the deep layer mean steering flow are the main source of the track errors for Hurricanes Harvey, Irma, and Maria (2017)
- Forecast errors of the Bermuda High dominated the steering flow errors for the 2017 hurricane season

Supporting Information:

Supporting Information may be found in the online version of this article.

Correspondence to:

T. W. Barbero,
tbarbero@colostate.edu

Citation:

Barbero, T. W., Bell, M. M., Chen, J.-H., & Klotzbach, P. J. (2024). A potential vorticity diagnosis of tropical cyclone track forecast errors. *Journal of Advances in Modeling Earth Systems*, *16*, e2023MS004008. https://doi.org/10.1029/2023MS004008

Received 7 SEP 2023
Accepted 25 FEB 2024

Author Contributions:

Conceptualization: Tyler W. Barbero, Jan-Huey Chen
Data curation: Jan-Huey Chen
Formal analysis: Tyler W. Barbero, Michael M. Bell, Jan-Huey Chen, Philip J. Klotzbach
Funding acquisition: Michael M. Bell, Jan-Huey Chen
Investigation: Tyler W. Barbero, Michael M. Bell, Jan-Huey Chen
Methodology: Tyler W. Barbero, Jan-Huey Chen

Abstract Tropical cyclone (TC) track forecasting provides essential guidance for coastal communities. However, track forecast errors still occur, highlighting the need for continued research into error sources. Piecewise potential vorticity (PV) inversion is used systematically to quantitatively diagnose errors in track forecasts in four models during the 2017 Atlantic hurricane season. The deep layer mean steering flow (DLMSF) provides a sufficient proxy for hurricane movement, and DLMSF errors are correlated with TC track errors. Analysis of track forecasts for Hurricanes Harvey, Irma, and Maria reveals that their track errors are attributed to steering errors caused by misrepresentations of specific pressure systems. Harvey's westward track error in the GFS resulted from zonal wind errors from the Continental High, while Irma's northward track error in the SHiELD gfsIC resulted from meridional wind errors in the Bermuda High and Continental High. Maria's southward track error in the IFS resulted from meridional wind errors in the Bermuda High and a misrepresentation of Jose to Maria's northwest. The mean absolute error of the DLMSF shows that the Bermuda High contributed the most to steering flow errors in the cases examined. Our results show that piecewise PV inversion can identify the sources of biases in TC track forecasts. The correction of these biases may lead to improved track forecasts. Quantitative diagnostics presented here provide useful information for future model development.

Plain Language Summary A tropical cyclone typically moves with the environmental wind, which is generated by several large-scale pressure systems (e.g., Bermuda High, Continental High) in the atmosphere. Weather models can predict the path of tropical cyclones, but these forecasts have errors. Tropical cyclones often bring devastation along their path, so it is important to mitigate track errors to provide better warnings for impacted communities. Here, we use a diagnostic technique called "piecewise potential vorticity inversion" to understand how the environmental wind causes errors in tropical cyclone tracks. In three different examples of hurricane track forecasts, we show that errors in the predicted track are caused by errors in the environmental wind from specific pressure systems. By considering numerous cases, we can also identify model biases, or errors that are consistent throughout many forecasts. These errors are a result of errors in the models themselves. Overall, our results show that piecewise potential vorticity inversion is a useful diagnostic tool that has the potential to improve track forecasts through the identification of model biases.

1. Introduction

Tropical cyclone (TC) track forecast errors have decreased substantially in the last several decades due to improvements in observational quality and quantity, numerical models and data assimilation techniques (Landsea & Cangialosi, 2018; Rappaport et al., 2009). Yet, small track forecast errors can still lead to large differences in impacts. For example, Hurricane Ian in 2022 made landfall near Fort Myers. If Ian had made landfall just 100 n mi north in Tampa as some weather model solutions forecasted, the impacts could have been even more devastating. Therefore, continued research into the sources of track error are imperative. Perfect track forecasts can never be achieved due to uncertainty in observations and modeling (Landsea & Cangialosi, 2018), but track errors can be further reduced through additional diagnostics that identify model errors and biases. In this study, we utilize a novel approach to diagnose the sources of TC track errors in numerical weather prediction models.

To a first order, TC movement generally follows the environmental wind. The environmental wind is usually associated with more than one large-scale atmospheric feature including subtropical highs, mid-latitude ridges and troughs, and low-level warm core systems (Marks, 1992). The mechanism by which the TC is steered by the mean flow is often known as the "cork in a stream" analogy. Quantitatively, TC movement can be approximated

Project administration: Michael M. Bell, Jan-Huey Chen
Resources: Michael M. Bell, Jan-Huey Chen, Philip J. Klotzbach
Software: Tyler W. Barbero, Jan-Huey Chen
Supervision: Michael M. Bell, Jan-Huey Chen, Philip J. Klotzbach
Validation: Tyler W. Barbero, Michael M. Bell, Jan-Huey Chen
Visualization: Tyler W. Barbero, Michael M. Bell, Jan-Huey Chen, Philip J. Klotzbach
Writing – review & editing: Tyler W. Barbero, Michael M. Bell, Jan-Huey Chen, Philip J. Klotzbach

by the response to the deep layer mean steering flow (DLMSF), which is the vertically-averaged horizontal wind in the vicinity of the TC (Chan & Gray, 1982). Both strength and proximity of the large-scale pressure system to the TC are critical to the contribution made to the DLMSF metric, but contributions from weaker systems that are farther away can also be important. From only the summation of these steering contributions, it is difficult to understand which systems contribute and how much each system contributes to the DLMSF. In this study, we employ the piecewise potential vorticity (PV) inversion technique (Davis, 1992a) to quantify and characterize the contribution of TC steering from individual large-scale pressure systems, as well as the error in the steering flow within model forecasts.

The application of PV to the atmosphere as a metric to understand the evolution of weather systems is described by the term "PV thinking" (Hoskins et al., 1985). Improved understanding of weather systems and their interaction is possible due to the properties of PV which are Lagrangian conservation and invertibility. Lagrangian conservation means that PV is conserved following adiabatic and inviscid motion, which is a reasonable assumption for large-scale atmospheric flow. Invertibility means that the thermodynamic and flow fields can be retrieved given a spatial distribution of PV and proper balanced condition as well as boundary conditions.

One application of PV thinking is piecewise PV inversion, where the PV is partitioned into "pieces," and each piece is inverted to deduce the global thermodynamic and wind fields pertaining to that specific PV distribution. In other words, piecewise PV inversion breaks down the total atmospheric flow into the main pressure systems that comprise it. This method has been applied to the investigation of several atmospheric processes including extratropical cyclogenesis and movement (Davis, 1992b; Davis & Emanuel, 1991; Hakim et al., 1996; Huo et al., 1999; Tamarin & Kaspi, 2016), TC movement (Shapiro, 1996; Wu & Emanuel, 1995a, 1995b; Wu & Kurihara, 1996; Wu et al., 2003, 2004, 2012), jet movement and development (Lackmann, 2002; Winters et al., 2020; Winters & Martin, 2017), precipitation forcing and processes (Baxter et al., 2011; Schlemmer et al., 2010; Wang et al., 2020), and frontogenesis (Chen et al., 2003; Korner & Martin, 2000; Morgan, 1999).

Several studies have used piecewise PV inversion to identify the causes of erroneous TC tracks in model forecasts. In Wu et al. (2004), the operational Aviation (AVN) model failed to predict the slowdown of Typhoon Sinlaku's (2002) westward motion and incorrectly forecasted a southward track of Sinlaku as it approached northeastern Taiwan. Using the piecewise PV inversion diagnosis, they found that the AVN model failed to predict the retreat of the western North Pacific subtropical high and building mid latitude trough. The steering from the western North Pacific subtropical high sustained Sinlaku's westward motion in the AVN forecast. The study also found that the overestimation of the Continental High over east Asia and underestimation of the western North Pacific Subtropical High in the AVN forecasts caused Sinlaku's southward track error. In Wu et al. (2012), the Global Forecast System (GFS) showed a more northwest track in Typhoon Sinlaku (2008) compared to its observed northeast track. The piecewise PV inversion diagnosis showed that the track difference was caused by a forecasted vortex height that was too shallow compared to the real height of the vortex. As a result, Sinlaku, in the GFS forecast, did not feel the contribution of upper-level westerly steering, and tracked northwestward instead of northeastward. The track forecast better aligned with the verified track after dropwindsonde data was assimilated into the GFS forecast. The vortex height was also better reproduced.

The aforementioned studies investigated track errors in cases of individual TCs, but none have used the piecewise PV inversion diagnosis on numerous TCs to identify consistently occurring errors (biases) in models. Model biases can be the result of errors in the model dynamics or physical parameterizations, data assimilation, initial conditions, or a combination of these. Several components work together to produce a model forecast, and it is difficult to attribute the exact cause of model biases. Furthermore, rectifying the model components producing the bias does not guarantee that new problems will not arise. Still, it is important to identify biases to be able to correct them and improve model forecasts.

Chen et al. (2019) identified a slow translation speed bias in the Integrated Forecast System (IFS) model run by the European Centre for Medium-Range Weather Forecasts (ECMWF) for Hurricanes Irma and Maria in 2017. This slow bias integrated over time and contributed to track errors. Following Chen et al. (2019), we use the piecewise PV inversion systematically on all TCs during the peak of the 2017 Atlantic hurricane season to identify the sources of model biases contributing to track errors in these storms for this season, such as the slow translation speed bias observed in the IFS.

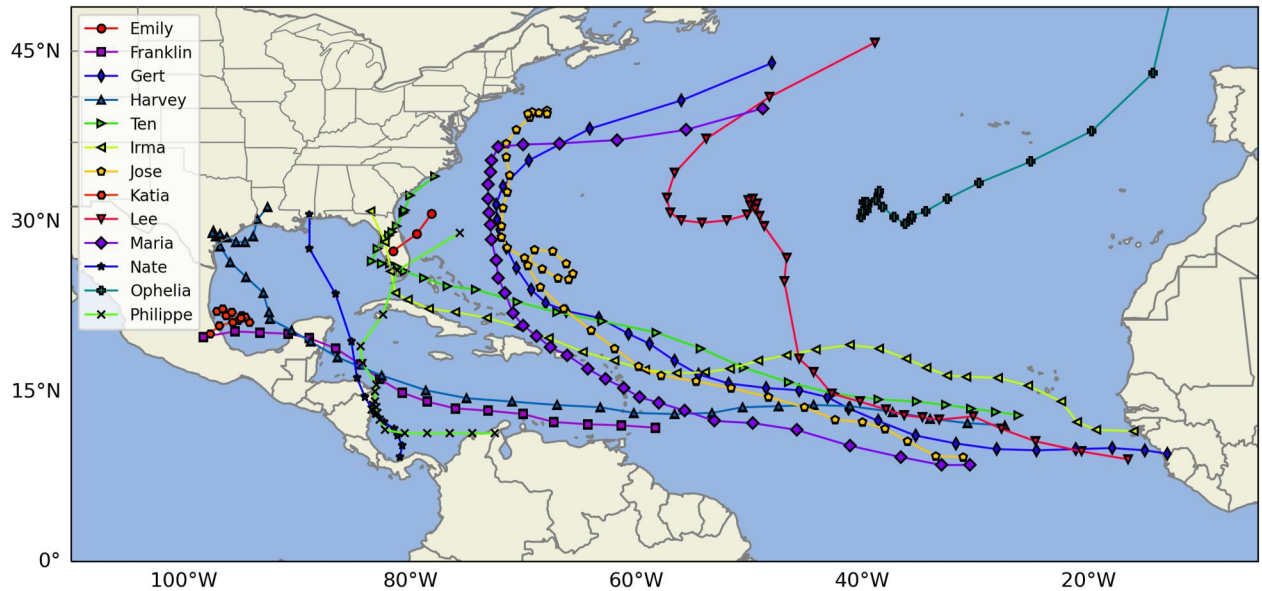

Figure 1. Best Track locations for the thirteen systems during August-October of the 2017 Atlantic hurricane season. Best Track locations are taken from the Automated Tropical Cyclone Best Track data set (Miller et al., 1990; Sampson & Schrader, 2000).

In this paper, an overview of the models, data, and PV framework is discussed in Section 2. In Section 3, three case studies of hurricane track forecast errors are analyzed to illustrate that the track errors are predominately caused by steering flow errors. Then, in Section 4, a basin-scale analysis of all TCs during the season was performed to investigate any biases stemming from the models within this season. Lastly, a brief summary and discussion of the limitations of this study are presented in Section 5.

2. Methodology

2.1. Data and Models

The 2017 North Atlantic hurricane season concluded with 6 major hurricanes (Category 3 and higher on the Saffir-Simpson Hurricane Wind Scale (Simpson, 1974)), 10 hurricanes, and 17 named storms with the 1981–2010 median number of major hurricanes, hurricanes, and named storms being 2, 6.5, and 12 respectively (Klotzbach et al., 2018). Figure 1 shows the tracks of the thirteen systems which occurred during August, September, and October of the 2017 Atlantic hurricane season.

The models utilized in this study include the National Centers for Environmental Prediction (NCEP) Global Forecast System (GFS; NCEP, 2016), the ECMWF Integrated Forecast System (IFS; ECMWF, 2016), and the formerly named fvGFS model, now called the System for High-resolution prediction on Earth-to-Local Domains (SHiELD; Harris et al., 2020) developed by the Geophysical Fluid Dynamics Laboratory (GFDL). The SHiELD model is able to ingest initial conditions from two prominent global prediction models: the GFS and IFS (Chen et al., 2018). Henceforth, the SHiELD model forecasts run with GFS initial conditions will be labeled SHiELD gfsIC and with IFS initial conditions will be labeled the SHiELD ifsIC. For a comprehensive overview of the model dynamical cores, physics suites, and native grid resolutions, please refer to Table 1 in Chen et al. (2019).

The native grid resolution and vertical levels are re-gridded to a 1° by 1° global latitude longitude grid and 25 vertical levels for all models. The 5-day forecasts are initialized every 12 hr from 1 August 2017 to 31 October 2017, and model outputs are saved at 6 hr intervals. The GFS and IFS analysis data at 00, 06, 12, and 18 UTC for the same time period are used to compare to forecasts. Diagnostics are performed over the region spanning 0° to 55°N and 110°W to 5°W. This grid encompasses the regions of interest including the North Atlantic and part of the continental United States.

TC track errors are computed by comparing the model forecast TC locations against their verified positions in the Automated Tropical Cyclone Forecast Best Track data set (Miller et al., 1990; Sampson & Schrader, 2000).

2.2. Potential Vorticity Framework

The dynamical framework used in this study follows Wu et al. (2003). Reformulated on the Exner function $\left(\pi = C_p\left(\frac{p}{p_o}\right)^{\kappa}\right)$ vertical coordinate and spherical horizontal coordinates, Ertel PV (Ertel, 1942; Rossby, 1940) can be defined as:

$$q = \frac{g\kappa\pi}{p}\left[\left(f + \nabla^2\Psi\right)\frac{\partial^2\Phi}{\partial\pi^2} - \frac{1}{a^2 cos^2\phi}\frac{\partial^2\Psi}{\partial\lambda\partial\pi}\frac{\partial^2\Phi}{\partial\lambda\partial\pi} - \frac{1}{a^2}\frac{\partial^2\Psi}{\partial\phi\partial\pi}\frac{\partial^2\Phi}{\partial\phi\partial\pi}\right], \tag{1}$$

where q is potential vorticity, Φ is the geopotential height, Ψ is the stream function for the non-divergent wind, $\kappa = \frac{R_d}{C_p}$, f is the Coriolis parameter, a is earth's radius, λ is longitude, ϕ is latitude, and the divergence and gradient operators are two-dimensional in the horizontal.

The Ertel PV can be inverted using the nonlinear balanced equation (Charney, 1955)

$$\nabla^2\Phi = \nabla\cdot(f\nabla\Psi) + \frac{2}{a^4\cos^2\phi}\frac{\partial(\partial\Psi/\partial\lambda, \partial\Psi/\partial\phi)}{\partial(\lambda,\phi)}. \tag{2}$$

Ertel PV is utilized rather than quasi-geostrophic PV (QGPV), because the accuracy of the QGPV inversion breaks down as the Rossby number approaches unity (Davis, 1992a). Large Rossby numbers may occur in highly rotational flows such as TCs. A number of studies have utilized the nonlinear balance equation to invert Ertel PV in studying TC movement and cyclogenesis (Davis & Emanuel, 1991; Shapiro, 1996; Wu & Emanuel, 1995a, 1995b; Wu et al., 2003, 2004).

The disadvantage of the framework is that the system being solved, Ertel PV and the nonlinear balance equation, is highly nonlinear. In other words, the sum of the solutions of piecewise PV perturbations does not add up to the total perturbation flow, which introduces ambiguity in the inversion interpretation. To compensate, the "full linear" method is applied which hides the nonlinear terms within a linear differential operator. This method effectively linearizes the system and removes interpretation ambiguity (Davis & Emanuel, 1991). A comprehensive comparison of linearization methods is reviewed in Davis (1992a).

In this study, we use the PV decomposition described in Shapiro (1996). The axisymmetric vortex in a 2,000 km radius around the TC center is calculated as the basic field, and perturbations are defined as deviations from the basic field. The basic state does not account for the background change in f. An f-plane is used such that the Beta effect is accounted for in the perturbation fields. Some of the perturbation will be due to variations in planetary rotation. The Shapiro decomposition is appropriate here because it cleanly separates the TC from its environment, allowing us to understand how the environment (i.e., PV perturbations) affects TC movement. It is important to note that the choice of decomposition is dependent on the atmospheric application.

The piecewise partitioning of the PV perturbation field (i.e., anomalies) is also highly dependent on the atmospheric feature of interest. For this study, we separate the PV anomalies geographically by sign (Figures 2a and 2b). All negative PV anomalies are divided by a meridional line at 80°W extending from 0° to 30°N connected to a diagonal line approximately parallel to the United States's East Coast. All positive PV anomalies are divided into four quadrants: northwest (NW), northeast (NE), southeast (SE), and southwest (SW), relative to the TC center at each point in time. The negative PV anomalies (i.e., Bermuda High and Continental High) correspond to large-scale pressure systems that are semi-permanent in space and time, and a static line dividing the Bermuda High and Continental High is appropriate. A simple division of the positive PV anomalies (i.e., low pressure systems) into four TC relative quadrants provides an efficient and practical way to attribute steering flow to these anomalies, instead of tracking them individually. Overall, the full PV perturbation field is partitioned into six regions where the sum of the pieces equals the total:

$$q'_{total} = q'_{BH-} + q'_{CH-} + q'_{NW+} + q'_{NE+} + q'_{\mathrm{SE}+} + q'_{SW+} \tag{3}$$

where BH is the Bermuda High and CH is the Continental High. The individual PV perturbations and the total PV perturbation are inverted individually to obtain the global, balanced flow fields corresponding to each

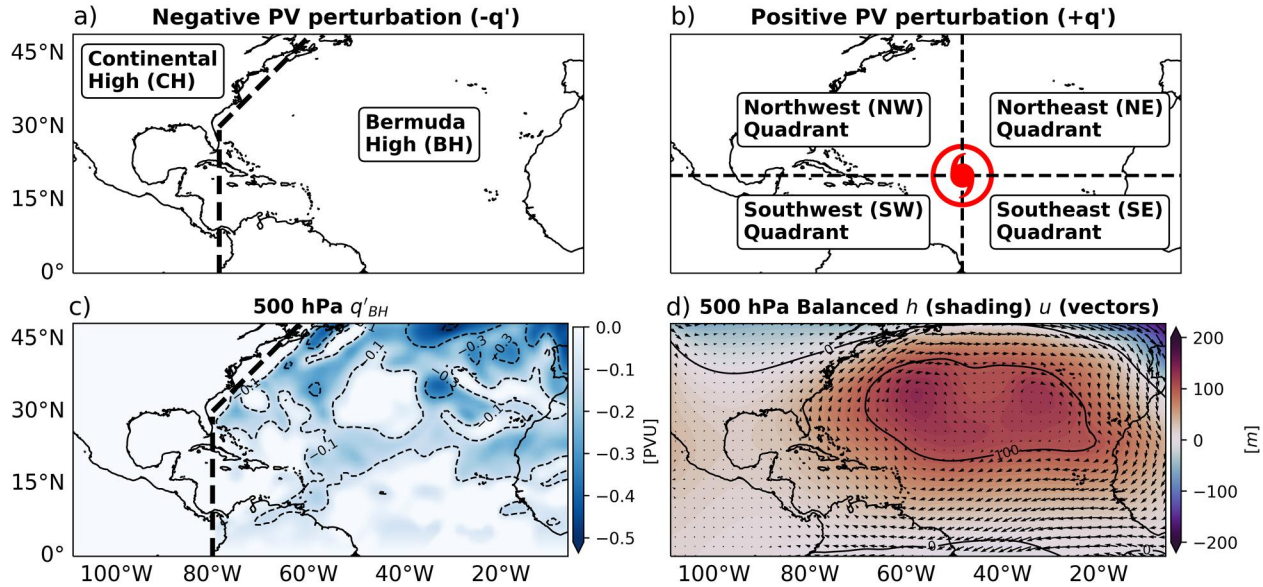

Figure 2. Partition of the total PV perturbation field (q') geographically into positive and negative anomalies. (a) Negative PV perturbations are divided into the Continental High and Bermuda High. A meridional line at 80°W extending from 0° to 30°N connected to a diagonal line approximately parallel to the United States's East Coast separates the PV systems. (b) Positive PV perturbations are divided into the northwest, northeast, southeast, and southwest quadrants relative to the center of the tropical cyclone at each point in time. (c) Example of partitioned Bermuda High q' at 500 hPa. (d) Global, balanced geopotential height (h) and wind (u) fields inverted from the Bermuda High q' at 500 hPa.

perturbation. To invert the PV, the geopotential height and stream function values are required on the lateral boundaries, and values of potential temperature are required on the top and bottom boundaries. For details of the inversion process, refer to Appendix A in Davis and Emanuel (1991).

The global, balanced flow fields approximately describe how each PV perturbation influences the atmosphere in the form of a geopotential height and stream function field. An example is shown in Figure 2. The negative PV perturbation at 500 hPa corresponding to the Bermuda High (Figure 2c) is isolated and inverted. As a result, the global, balanced flow fields associated with the Bermuda High are recovered (Figure 2d). From the stream function, the non-divergent horizontal wind field can be computed at each atmospheric level from $V = \hat{k} \times \nabla\Psi$, which is what is plotted in Figure 2d. The DLMSF, which is used to represent TC movement in this study, is defined as the vertically-averaged, balanced, non-divergent, horizontal wind from 925 to 300 hPa in a 3° radius around the TC center. The DLMSF is calculated for each of the balanced wind field regions for the entire forecast time.

2.3. Bridging Track Error and Steering Flow Error

TC movement to first order is driven by the environmental wind, but the interaction of the TC with its environment (i.e., beta drift) can cause the TC to deviate from the large-scale steering flow, typically poleward and westward. Here, we consider movement only by the large-scale environment which assumes that errors in the track forecast are due solely to errors in the wind field surrounding the TC. Therefore, we use DLMSF error as a proxy for track error. The correlation between the DLMSF error at varying lead times and track errors are shown in Figure 3 at different static lead times. All of the models offer 6 hourly output to 5 days, but the correlation was calculated using only homogeneous cases in which forecasts initialized at a particular time had a TC in every model. For example, a model may not have a TC present at a particular forecast lead time because it predicted the TC to dissipate. These forecast periods are removed from the homogeneous set.

The DLMSF error and track error correlations are moderately high with a Pearson correlation coefficient of 0.5–0.75 for the 72-, 96-, and 120-hr static forecast lead times and lower correlations of around 0.25–0.50 for the 48-hr static forecast time. The correlation plots generally show a maximum at −24 to 0 hr, meaning track errors at a certain lead time are generally caused by errors in the TC DLMSF up to 24 hr before. The time lag for maximum correlation is reasonable, as it takes time for steering errors to accumulate and cause errors in the position at later

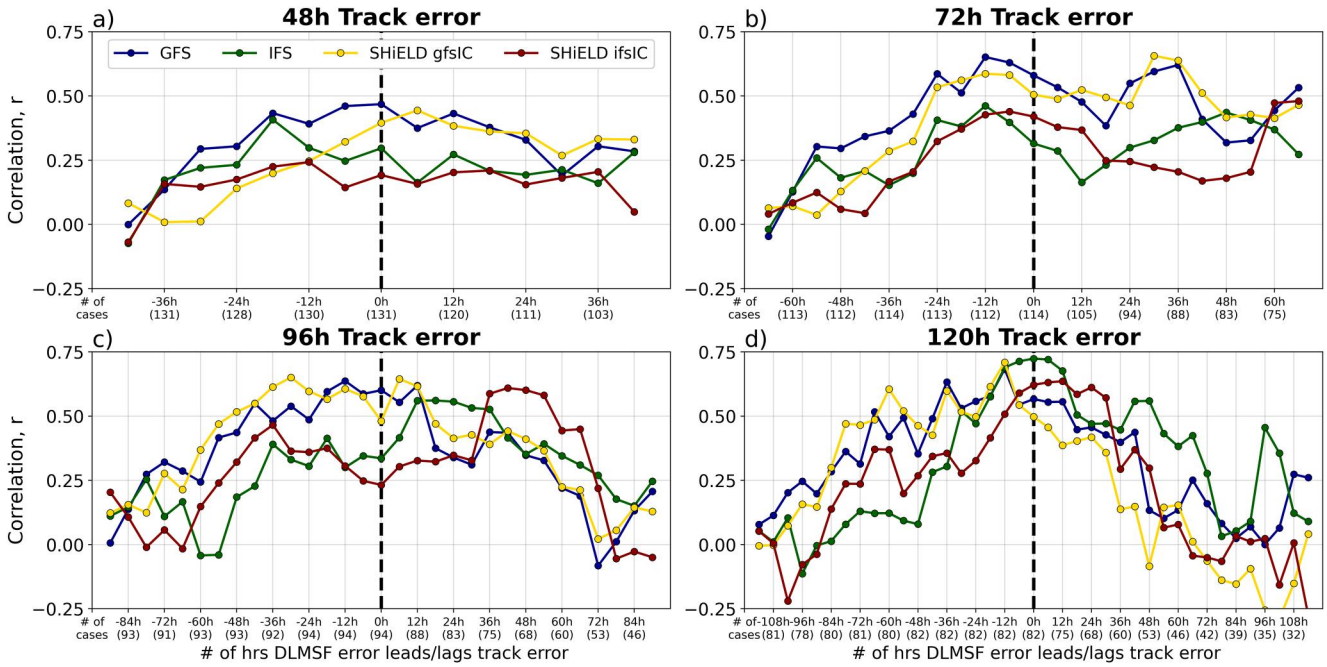

Figure 3. Correlations between the DLMSF error and TC track error in model forecasts. The DLMSF error varies relative to the track error at a static forecast hour. The correlations are shown for different static forecast lead times: (a) 48 hr, (b) 72 hr, (c) 96 hr, and (d) 120 hr. The number of cases shown along the x-axis are the homogeneous cases.

forecast times. Interestingly, when DLMSF error lags track error (e.g., positive hours), the correlation does not immediately decrease to zero. A probable reason is that storms that are forecast in the wrong place (i.e., track error) will possibly lead to errors in the steering flow (i.e., DLMSF errors), and therefore track errors leading DLMSF errors will be lag correlated.

The 48 hr correlation plot shows the lowest correlation values out of all the lead times. The low correlations may be due to the high performance of models at earlier lead times, which means that the DLMSF errors and track errors are low and have not had sufficient time to cause errors in the other fields. At later lead times, errors may have had time to develop and contribute errors to other fields. This would explain the increasing maximum correlation with increasing lead time.

While the correlations between track errors and DLMSF errors are relatively high, the correlations are not perfect, meaning that the DLMSF error is an imperfect proxy for track error. The discrepancy between the two could be due to the representation of the DLMSF and how our methodology only considers advection as well as simplifications in the PV inversion methodology. We elaborate more on the DLMSF limitations in Section 5. Overall, the moderate correlations between the DLMSF error and the track error show that our defined DLMSF is a sufficient proxy for the forecasted movement of a TC, and that DLMSF errors are representative of track errors.

3. Case Studies Analyses

The utility of the piecewise PV inversion diagnostic is now shown for three case studies of hurricanes (Harvey, Irma, and Maria) within different steering environments. For each case, a 5-day track forecast exhibiting distinctive track errors is chosen and analyzed. We present the track forecasts for all four models along with the Best Track in each case study. For brevity, we only analyze two models for each case in detail to examine both a good and poor model track forecast.

3.1. Case Study: Hurricane Harvey

At 03 UTC on 26 August 2017, Hurricane Harvey made landfall at Category 4 intensity (Saffir-Simpson Hurricane Wind scale) along the central Texas coastline. As Harvey approached the coastline of Texas, its northward movement slowed to an eventual stall. Harvey's proximity to the Gulf of Mexico at the time of its stall provided

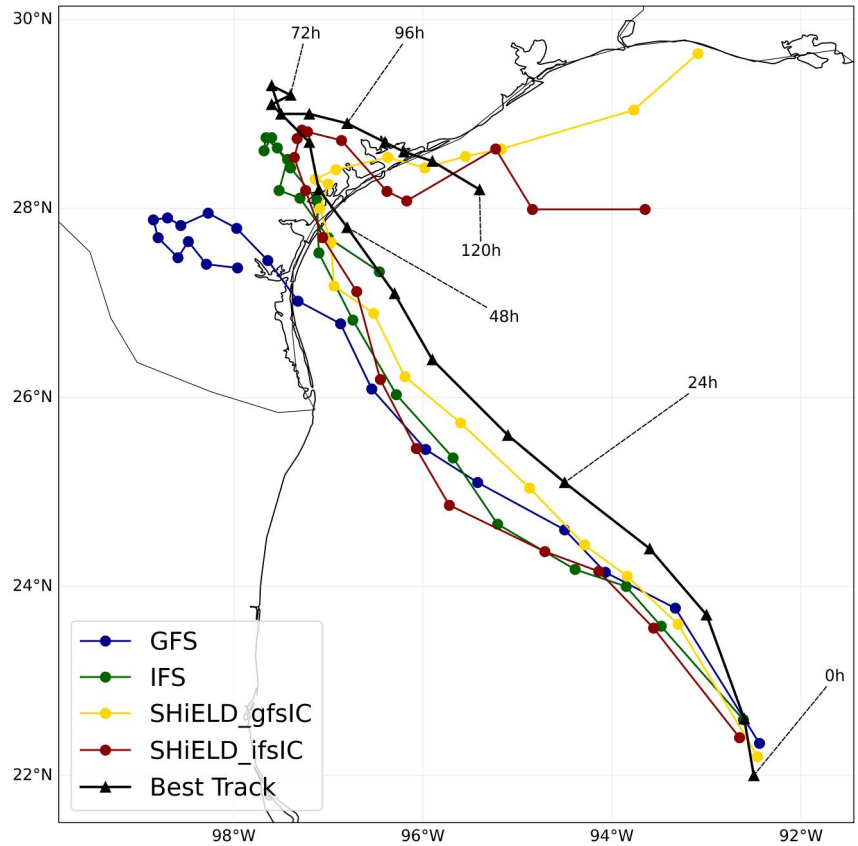

Figure 4. 5-day track forecast and best track positions for Hurricane Harvey. The track forecast positions start at 6 hr following the initialization time and are plotted every 6 hr. The best track positions start on 00 UTC 24 August 2017 and are plotted every 6 hr.

optimal conditions for moisture transport from the Gulf of Mexico over land, resulting in several feet of rainfall in population-dense regions of southeastern Texas including the Houston metropolitan area (Blake & Zelinsky, 2018; DeHart & Bell, 2020). However, the forecasted track of Harvey by various models exhibited errors in both speed and direction. These differences in northward translation speed or landfall position would likely have resulted in vastly different societal impacts.

Figure 4 shows Harvey's 5-day track forecast initialized at 00 UTC 24 August 2017 with 6 hourly forecasts and coinciding best track (BT) positions. By comparing the model track forecasts to the BT, we see that none of the models forecasted every aspect of Harvey's track correctly, but they all exhibited a stall as Harvey traversed onto land. While the IFS and SHiELD models moved in approximately the correct direction, the GFS stalled much farther west than the other models and the BT. The piecewise PV inversion tool is used to diagnose the westward track error in the GFS forecast. The GFS forecast is compared to the SHiELD gfsIC forecast, which performed relatively well. The GFS analysis provides a baseline with which to compare the model forecasts.

Figure 5a shows the balanced, geopotential heights at 500 hPa and balanced mean horizontal wind in the 925–300 hPa layer in the GFS analysis. The demarcation between the Continental High and Bermuda High PV is included for clarity, but the balanced fields shown are from the full perturbation PV field. The analyzed synoptic setup around Harvey at 00 UTC 27 August 2017 in Figure 5a showed that Harvey was located between two high pressure systems: a high located over the Four Corners region to the northwest, and a high pressure system located over the North Atlantic subtropics to the east. The two highs created opposing, anticyclonic circulations, resulting in a weak steering environment in Harvey's near-environment flow. A weak low pressure system to the southwest is also discernible with associated weak cyclonic flow, but this low pressure system was not readily apparent in the mid-level geopotential height field. Another low pressure system was located over the northeastern U.S., but this system was likely too distant to play a major role in Harvey's steering.

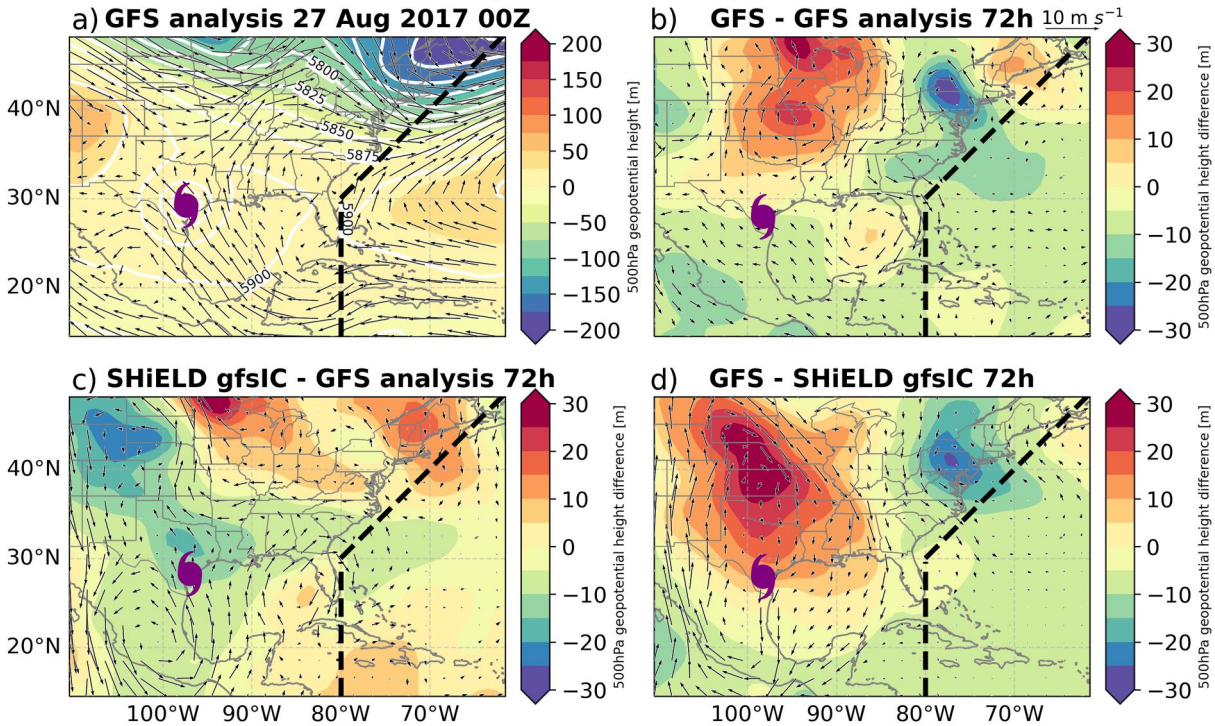

Figure 5. Overview of the synoptic features surrounding Harvey at the 72 hr lead time. 500 hPa balanced geopotential heights (m) are shaded. 925–300 hPa layer-mean balanced wind fields are illustrated by the arrows (m s^{-1}). The thick, dashed line represents the static boundary separating the Continental High and Bermuda High regions. Panel (a) shows the GFS analysis fields as well as the GFS analysis 500 hPa total geopotential height field overlaid (contours) at 27 August 2017 00 UTC concurrent to the time of the 72 hr lead time forecasts. Panels (b)–(d) show the differenced fields for the GFS forecast−GFS analysis, the SHiELD gfsIC−GFS analysis, and the GFS forecast−SHiELD gfsIC forecast, respectively.

Errors in the strength and location of these large-scale systems likely resulted in forecast track errors. An overview of the errors in the steering flow surrounding Harvey is shown by the differences of these two fields between the GFS and the GFS analysis at 72hr lead time in Figure 5b. The main factor driving Harvey's westward track error is the presence of higher heights and stronger anticyclonic flow associated with the Continental High to the north. This flow blocked Harvey's path northward and instead steered Harvey toward the west. Lower heights and stronger cyclonic flow associated with a low to the south also steered Harvey to the west. However, the influence from the low likely did not play as large of a part in Harvey's westward track error since the low was relatively weak. The SHiELD gfsIC forecast error is shown in Figure 5c. The forecast shows weaker heights collocated with stronger cyclonic flow compared to the analysis, but the height differences are small relative to the GFS forecast error. Consequently, the small difference in steering flow in the SHiELD gfsIC allowed for a track forecast that better follows the BT. The differences between the GFS and SHiELD gfsIC forecasts are shown in Figure 5d. The high to the north of Harvey in the GFS forecast is much stronger than in the SHiELD gfsIC forecast, which contributed easterly steering over Harvey consistent with the westward track error shown in the GFS.

Figure 6 shows the DLMSF error contributions from each partitioned system for the GFS forecast. For brevity, only the DLMSF error from the erroneous GFS forecast is shown. At the top of each subplot is the observed storm motion which is shown by the blue vectors and the GFS forecast storm motion which is shown by the red vectors. Harvey's westward track deviation in the GFS forecast can be represented by the total contribution of the DLMSF errors from the partitioned systems which generally starts around 48–72hr lead time. The Continental High produces the largest DLMSF error to the north-northwest. Moderate DLMSF error to the west-northwest is contributed by the low quadrants. Lower DLMSF error to the west is contributed by the Bermuda High. Overall the largest source of error is from the Continental High producing up to 3 m s^{-1}, with moderate error around 0.5 m s^{-1} from the low quadrants, and a smaller error from the Bermuda High, generally less than 0.5 m s^{-1}.

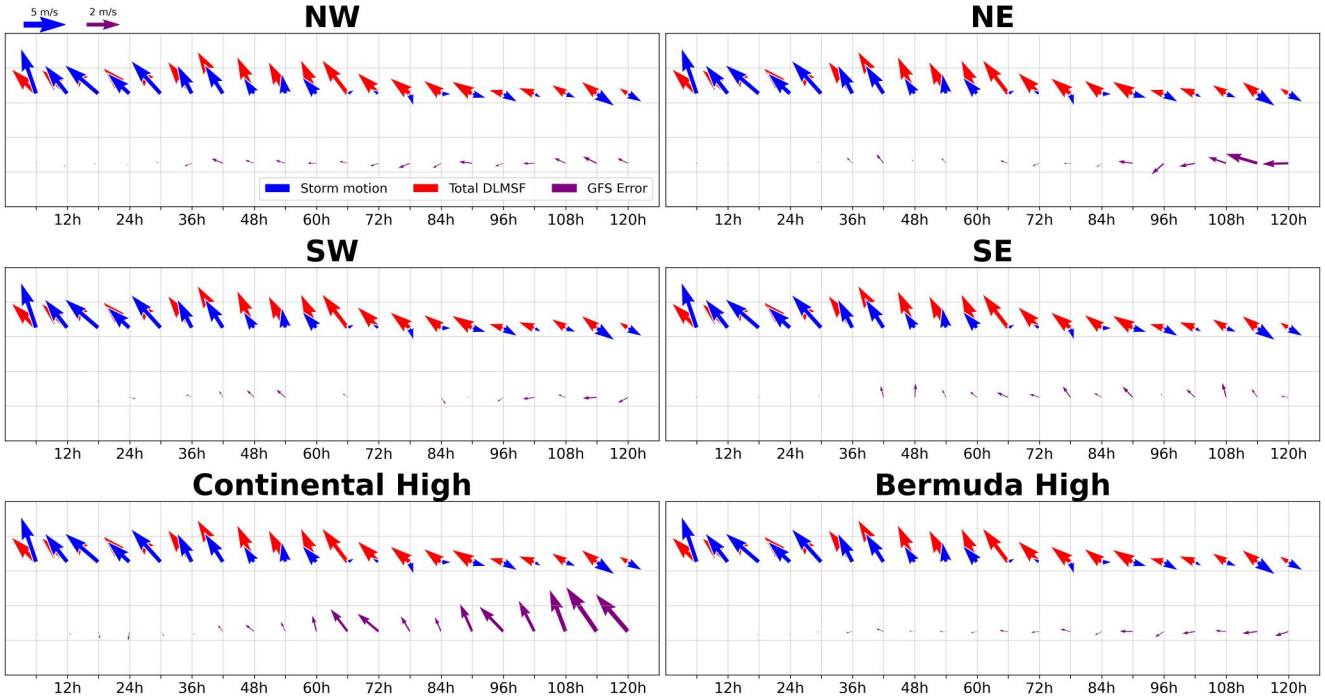

Figure 6. Plots of the DLMSF error (purple vectors, m s^{-1}) for Harvey in the GFS. Panel (a)–(f) represents the DLMSF error from the NW quadrant, NE quadrant, SW quadrant, SE quadrant, Continental High, and the Bermuda High, respectively. The observed storm motion (blue vectors, m s^{-1}) and GFS storm motion (red vectors, m s^{-1}) are also plotted at the top of each subplot.

3.2. Case Study: Hurricane Irma

Hurricane Irma was a powerful Category 5 hurricane that caused devastation to several Caribbean islands including Cuba and then portions of Florida. Its track was governed by a combination of steering from the Bermuda High, Continental High, a distant, mid-latitude trough to the northeast, and a developing low to its northwest (Cangialosi et al., 2021). Like Harvey, differences in Irma's forecast track would have caused Irma to have drastically different impacts.

Figure 7 shows the 5-day track forecast for Irma every 6 hr initialized at 00 UTC 8 September 2017 along with the corresponding BT positions. By 6 hr lead time, the forecasts began to diverge from each other, although models run with the same initial conditions had similar tracks. The models run with GFS initial conditions showed a northward track error as well as a slight slow translation speed error, while the models run with IFS initial conditions generally followed the BT. The northward track error and slow translation speed error of Irma in the SHiELD gfsIC is now compared to the relatively well forecasted track in the SHiELD ifsIC.

The synoptic-scale features influencing Irma's movement at 24hr lead time are shown in Figures 8a and 8b. The analysis fields show three main systems: a high pressure to the east, a high pressure to the west, and a trough to the north. Steering flow in the near surroundings of Irma was largely southeasterly, dominated by the influence of the Bermuda High to its east.

The SHiELD gfsIC forecast error is shown in Figure 8c. A dipole of weaker heights to the west and higher heights to the east was collocated with stronger southerly steering flow, which supports the northward track error in the SHiELD gfsIC. The SHiELD ifsIC forecast error in Figure 8d also shows differences in the height field, but these differences are smaller compared to the SHiELD gfsIC. These smaller differences in the SHiELD ifsIC do not manifest into a northward track error as the SHiELD gfsIC does.

The DLMSF errors from each system for the SHiELD gfsIC forecast are shown in Figure 9. Irma's observed storm motion and its storm motion in the SHiELD gfsIC forecast are also shown at the top of each subplot. The error vectors for Irma and Harvey differ, with Harvey generally having larger errors. Consequently, the errors in Harvey are not visually comparable to those in Irma. Irma's northward track deviation in the SHiELD gfsIC

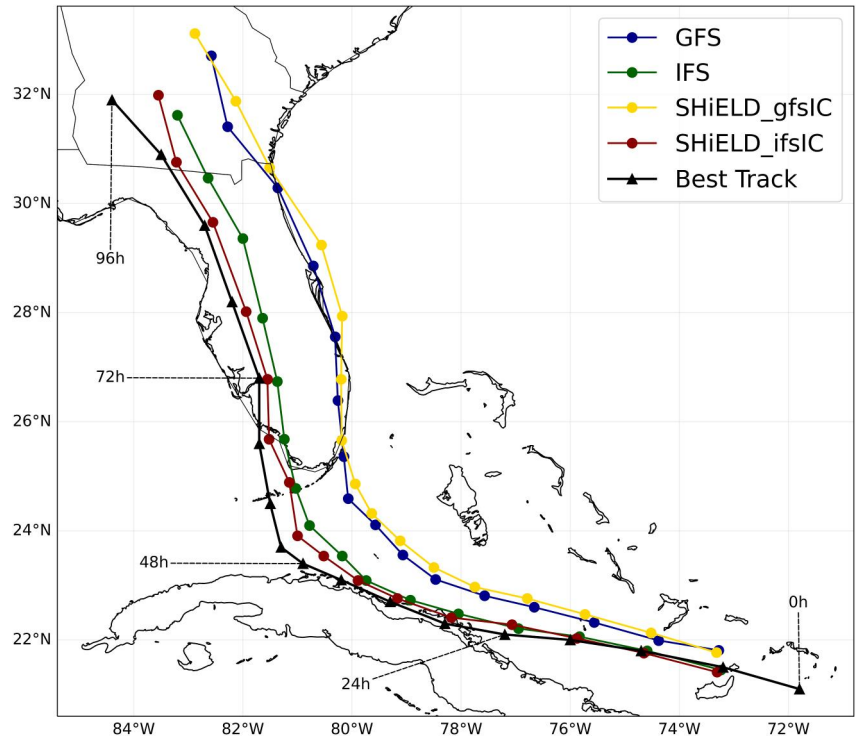

Figure 7. As in Figure 4 but for Hurricane Irma. The track forecast was initialized from 00 UTC 8 September 2017. The tracks stop at 96 hr due to Irma being declared a remnant low at that time.

forecast can be explained by the DLMSF error from the partitioned systems. We see that the Bermuda High and Continental High produce the largest DLMSF errors, with the low quadrants contributing smaller errors. The direction of DLMSF error contributed by the Bermuda High and the Continental High oppose each other, with the Bermuda High mainly producing a northward error and the Continental High mainly producing a southward error on Irma. While the two systems exhibit comparable errors, northward flow from the Bermuda High is larger in magnitude by 1–2 m s^{-1} than the southward flow from the Continental High. This resulted in a net northward contribution to the total DLMSF, consistent with the northward track bias shown in the SHiELD gfsIC.

Irma's slight slow translation speed bias in the SHiELD gfsIC forecast is illustrated by the DLMSF error from the partitioned systems. Eastward DLMSF error from the Bermuda High, Continental High and NE quadrant are produced while westward DLMSF errors are produced by the NW, SW, and SE quadrants at various lead times. However, the largest contributors to the DLMSF error are the Bermuda High and the Continental High which each contributed around 0.5 m s^{-1} DLMSF error, whereas the low quadrants contribute around 0.25 m s^{-1} error. Overall, the slight slow translation speed bias is due to a combination of eastward DLMSF error from different systems at various lead times.

3.3. Case Study: Hurricane Maria

Hurricane Maria was a devastating hurricane that made landfall in Dominica as a Category 5 hurricane, followed by a second landfall in Puerto Rico at Category 4 intensity. Maria also posed a significant track forecasting challenge. Its track motion was governed mainly by the Bermuda High (Pasch et al., 2023).

The 5-day track forecast was initialized at 00 UTC 17 September 2017 and is shown in Figure 10 along with the BT positions for the same time. The models ingesting GFS initial conditions show a northward track error at early lead times continuing through most of the forecast. The models initialized with IFS initial conditions follow a path similar to the BT. However, the IFS showed a southward track error beginning around 48hr lead time, which caused the IFS forecast to narrowly miss making landfall in Puerto Rico. Interestingly, the IFS also showed a slow translation speed error around the same time. The southward track error in the IFS will be analyzed here.

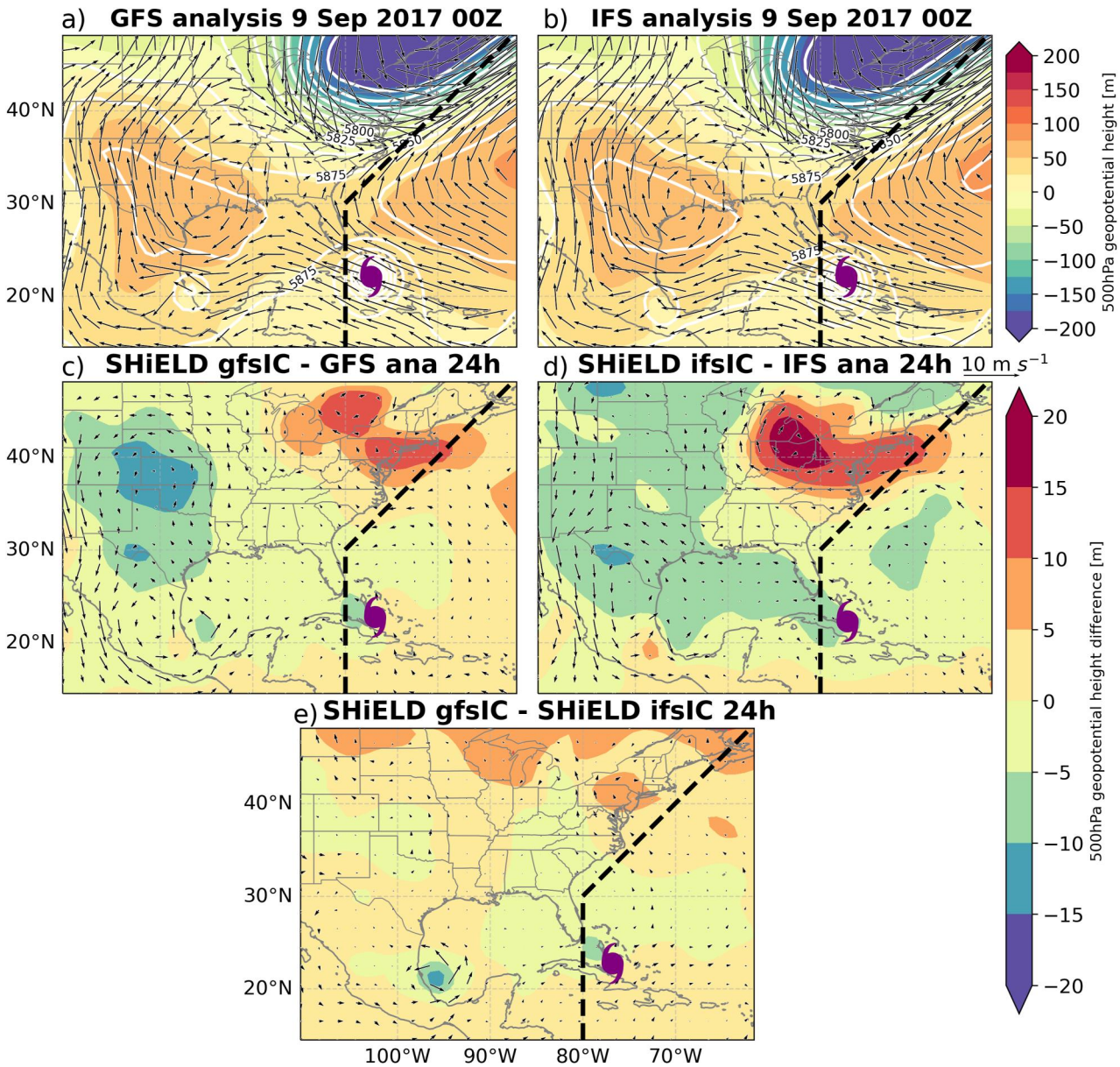

Figure 8. Overview of the synoptic features surrounding Irma at the 24 hr lead time. 500 hPa balanced geopotential heights (m) are shaded. 925–300 hPa layer-mean balanced wind fields are illustrated by the arrows (m s^{-1}). Here, the balanced fields are retrieved from the full PV perturbation field. The thick, dashed line represents the static boundary separating the Continental High and Bermuda High regions. Panel (a) shows the GFS analysis balanced heights and layer-mean wind as well as the 500 hPa total geopotential height field overlaid (contours). Panel (b) shows the same as in (a) but for the IFS analysis fields. Panels (c)–(e) show the differenced fields for the SHiELD gfsIC forecast−GFS analysis, the SHiELD ifsIC−IFS analysis, and the SHiELD gfsIC forecast−SHiELD ifsIC forecast, respectively.

An overview of the environmental, synoptic features is shown in Figures 11a and 11b. The analysis fields show four main pressure systems of interest: a strong low pressure to the northwest (i.e., Hurricane Jose), high pressure to the north, a trough to the northeast, and the Bermuda High to the east. Maria is embedded within strong southeasterly winds imposed by the high pressure systems. The IFS forecast error (Figure 11d) showed a weaker high to the east and north, which imparts stronger northerly flow on Maria, potentially explaining its southward track error at 72hr lead time. Interestingly, a strong dipole of height differences to the northwest was due to differences in position and intensity of Hurricane Jose in the IFS forecast compared to its analysis field. The GFS forecast error (Figure 11c) was relatively low, thus Maria's motion to the northwest in the GFS did not deviate very much. The model forecast difference is shown in Figure 11e, showing that the IFS exhibited more north-easterly steering on Maria, reinforcing its southward track error.

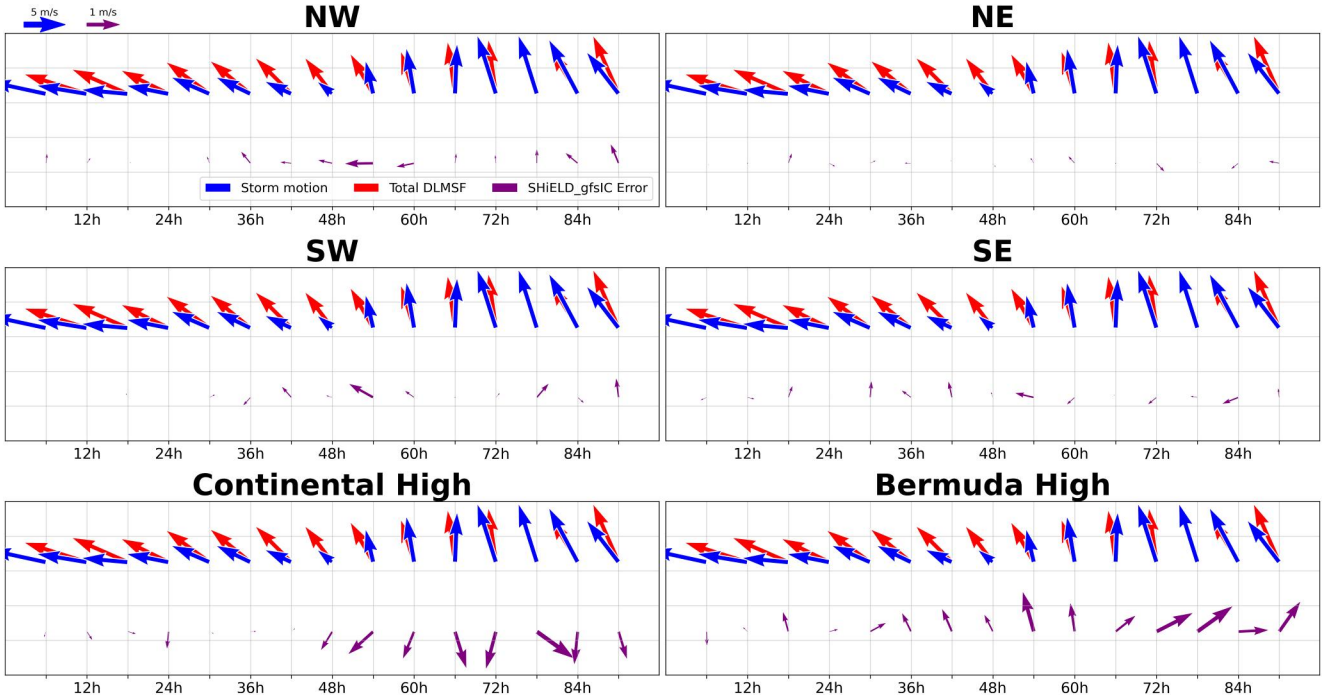

Figure 9. As in Figure 6 but for Hurricane Irma and the SHiELD gfsIC. Note that the error vector reference value for Irma is 1 m s^{-1} compared to 2 m s^{-1} for Harvey.

Figure 12 quantifies the DLMSF error from the individual systems on Maria for the IFS forecast. Unlike Harvey, the error vectors for Maria and Irma share the same scale and can be visually compared. This case is not as straightforward to interpret as the previous two cases. The IFS exhibits a southward track error at around 72 hr lead time. The Bermuda High, NW and SW quadrants also show strong southward DLMSF error on Maria in the IFS which supports its southward track error. However, the Continental High, and SE and NE quadrants also exhibit strong northward DLMSF error which does not support the southward track error. A further look shows that the systems produce varying amounts of meridional steering with the strongest steering flow from the Bermuda High at 3–4 m s^{-1}, the Continental High and NE quadrant at 2–3 m s^{-1}, and smaller contributions of 1–2 m s^{-1} from the NW, SE, and SW quadrants. Overall, the contribution from the southward steering flows is slightly larger than that of the northward flow errors, resulting in a net southward track.

Recall that the IFS forecast showed large errors from Hurricane Jose (Figure 11d). While large errors in geopotential heights were evident to the northwest of Maria, the signal is not obvious in the DLMSF error. Because Hurricane Jose existed on the boundaries between the Bermuda High and Continental High, and NE and NW

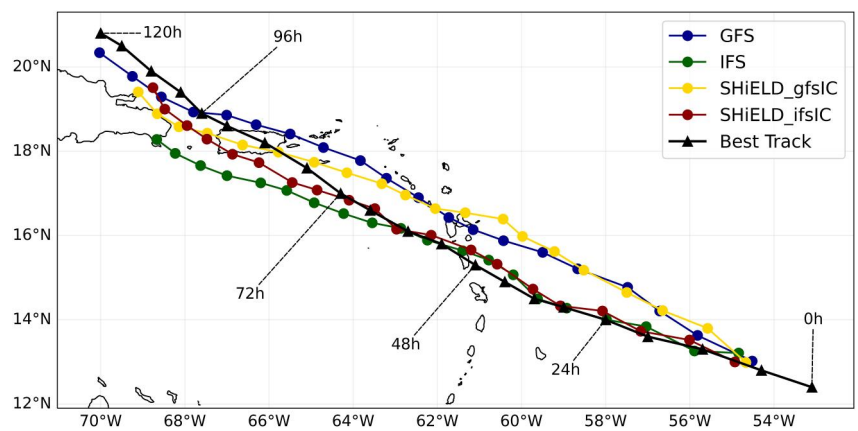

Figure 10. As in Figure 4 but for Hurricane Maria. The track forecasts are initialized at 00 UTC 17 September 2017.

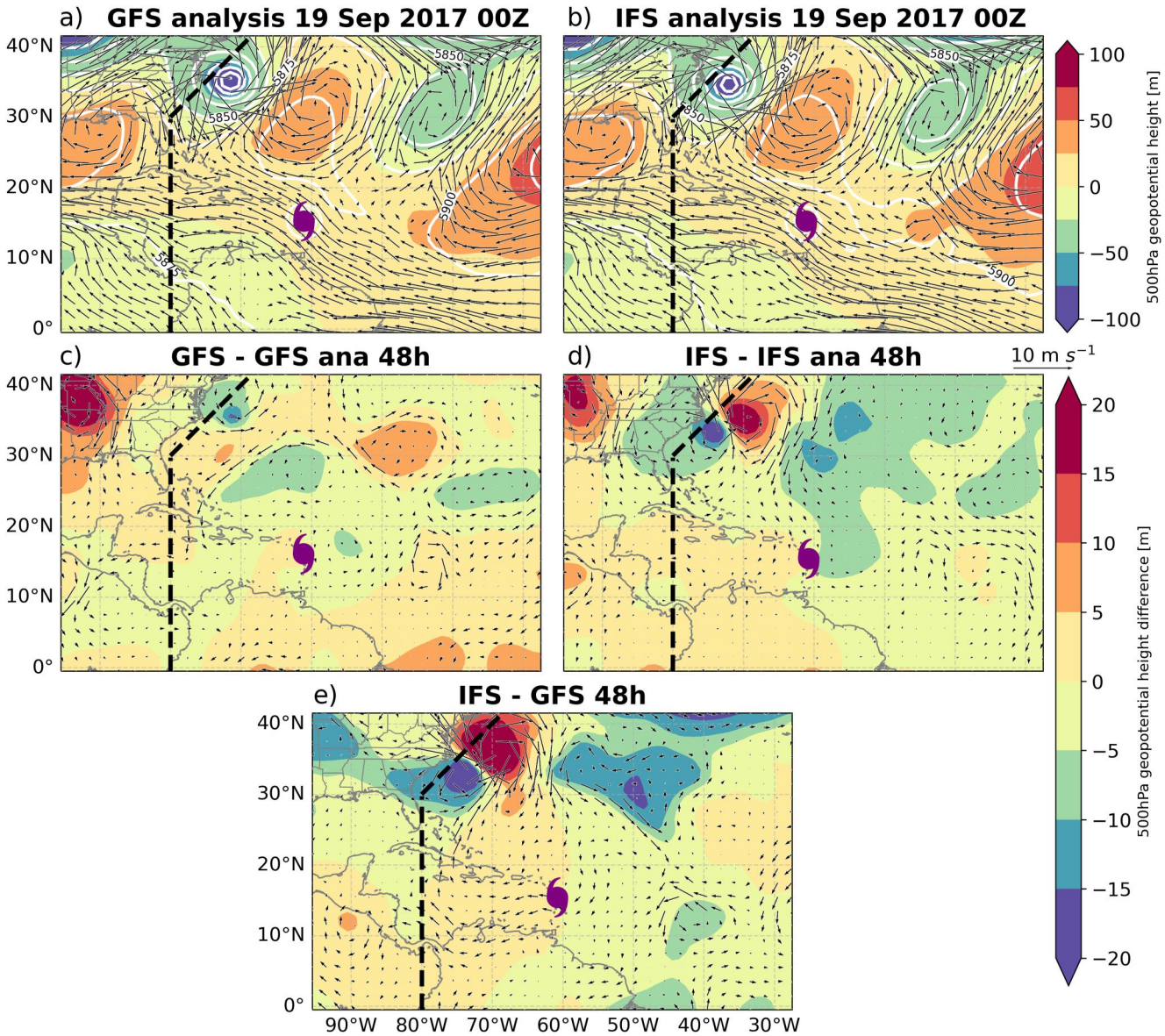

Figure 11. Overview of the synoptic features surrounding Maria at the 48 hr lead time. 500 hPa balanced geopotential heights (m) are shaded. 925–300 hPa layer-mean balanced wind fields are illustrated by the arrows (m s^{-1}). Here, the balanced fields are retrieved from the full PV perturbation field. The thick, dashed line represents the static boundary separating the Continental High and Bermuda High regions. Panel (a) shows the GFS analysis balanced heights and layer-mean wind as well as the 500 hPa total geopotential height field overlaid (contours). Panel (b) shows the same as in (a) but for the IFS analysis fields. Panels (c)–(e) show the differenced fields for the GFS forecast−GFS analysis, the IFS forecast−IFS analysis, and the IFS forecast−GFS forecast, respectively.

quadrants, the signal in the DLMSF error was distributed between them. This is a drawback of our PV partitioning scheme where the interpretability of steering flows that cause track errors can become ambiguous.

4. Basin-Scale Analyses

4.1. DLMSF Error Statistics

The piecewise PV inversion tool was performed on 13 systems during the 2017 North Atlantic hurricane season including Emily, Franklin, Gert, Harvey, Potential Tropical Cyclone Ten, Irma, Jose, Katia, Lee, Maria, Nate, Ophelia, and Philippe. A common metric to quantify the amount of error in a prediction is the mean absolute error (MAE). The DLMSF MAE is calculated by summing the absolute value of the DLMSF error and dividing by the number of homogeneous cases at a particular lead time. Figure 13 quantifies the average DLMSF attributed to the

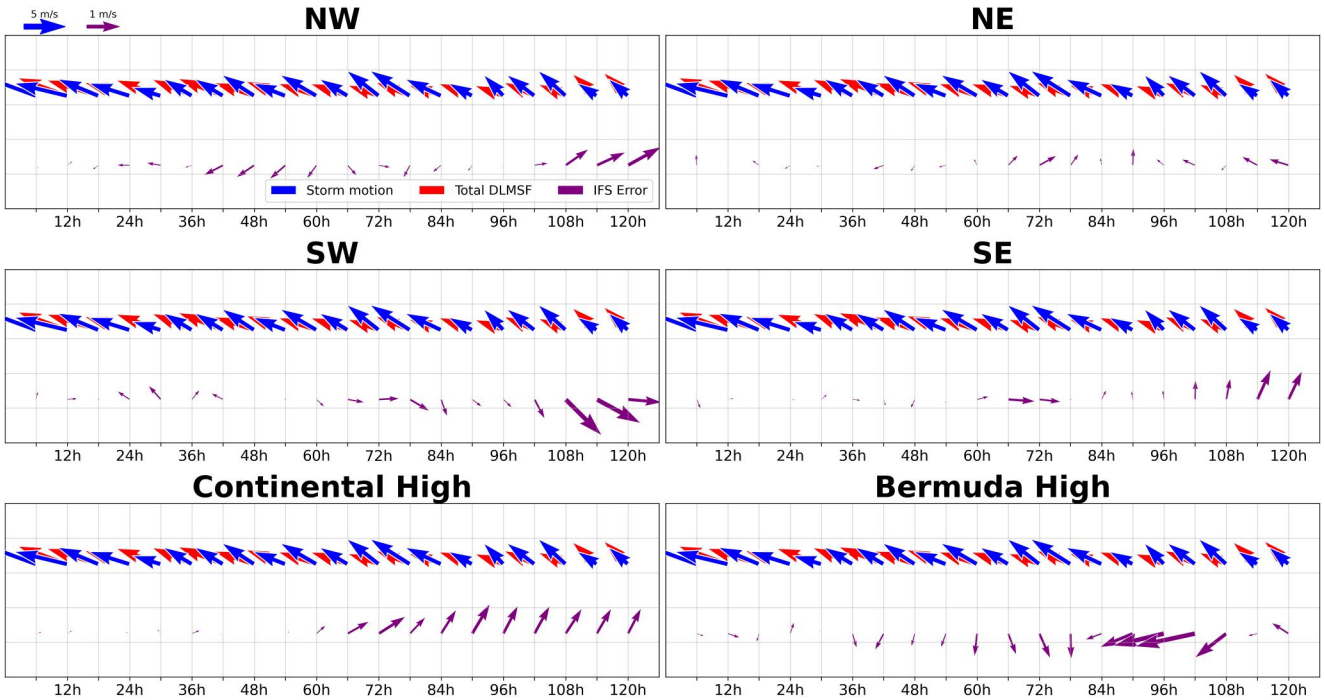

Figure 12. As in Figure 6 but for Hurricane Maria and the IFS. Note that the error vector reference value for Maria is 1 m s^{-1} compared to 2 m s^{-1} in Harvey.

different PV regions as a function of model and lead time. To a first order, the DLMSF MAE increases in time for all models and systems as expected. The Bermuda High contributes the largest DLMSF MAE, especially at later lead times. The Continental High produces a moderate amount of error, and the storm relative quadrants produce smaller errors. Generally, the DLMSF MAE increases linearly through 120 hr lead time.

While the MAE describes the absolute average deviation of a measurement from the mean, it is important to also quantify the sign of the deviation (bias). Figure 14 illustrates the bias for the partitioned systems for every model. A positive mean (i.e., positive bias) indicates that the forecasted DLMSF is stronger than the analysis, and a negative mean (i.e., negative bias) shows that the forecasted DLMSF is weaker than the analysis. A mean near zero indicates that the DLMSF forecast errors are evenly distributed between positive and negative errors, resulting in a near-neutral bias. The ±1 sigma bounds are also shaded for each model over the 5-day forecast.

The strongest biases are present in the Bermuda High which suggests that the Bermuda High is a significant producer of TC movement errors. The DLMSF error bias from the Bermuda High in the IFS forecasts shows a strong consistent negative bias, and indicates that the steering flow from this region is simulated too weakly. In contrast, the SHiELD model forecasts produce a slight positive bias which means the DLMSF from the Bermuda High is slightly stronger than the observed steering flow. On the other hand, the GFS forecasts show a near zero DLMSF error bias until later lead times, where the bias becomes slightly negative.

Moderate biases are present in the Continental High. The SHiELD model forecasts show a negative bias that transitions to a near neutral bias, indicating that these models tend to underpredict the strength of steering flow from the Continental High at earlier lead times. This underprediction evens out at later lead times. The IFS forecasts show a near neutral bias for most lead times. The GFS forecast biases in the Continental High are mainly positive.

The four quadrants illustrate smaller, near-neutral biases compared to the Bermuda High and Continental High. A reason for the near zero bias for the quadrants is that the quadrants are storm-relative, and the biases may be averaged out. Another way to visualize the bias in the DLMSF error is through probability density functions, which are included in Figures S1–S4 in Supporting Information S1 and show the DLMSF error distributions for the GFS, IFS, SHiELD gfsIC, and SHiELD ifsIC, respectively.

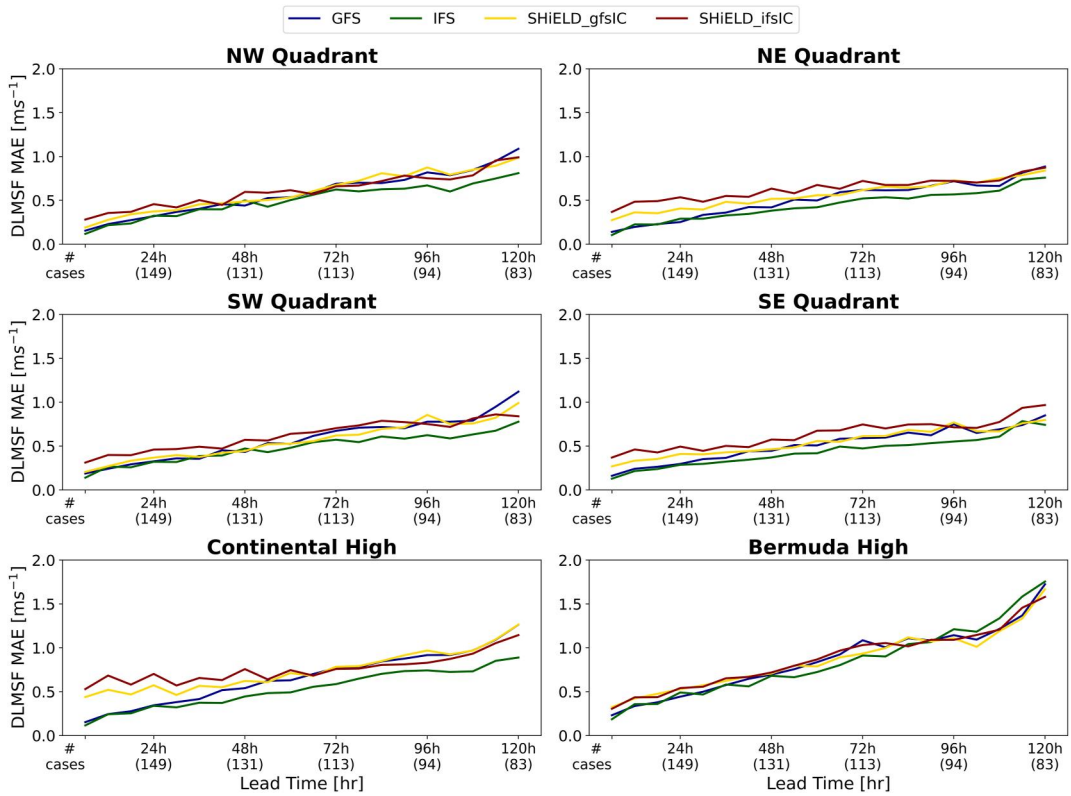

Figure 13. Mean absolute error (MAE) of the DLMSF as a function of forecast hour. The number of cases shown along the *x*-axis are the homogeneous cases.

The standard deviation of the DLMSF errors is plotted in Figure 15. The standard deviation is calculated similarly to the MAE, and both metrics illustrate the deviation of errors from the mean error. However, the standard deviation is the square root of the square of the errors compared to the absolute value of the error in the MAE, such that larger error differences are weighted more. Thus, standard deviation values are typically larger than MAE values. As expected, the standard deviation of the DLMSF errors follows very similarly to the MAE in Figure 13.

4.2. Composite Analyses

As shown previously, the Bermuda High has the largest DLMSF MAE and contributes the most to TC movement error. This large MAE is more apparent at later lead times. To understand the result of the high DLMSF MAE, a composite analysis of the balanced geopotential height at 500 hPa associated with the Bermuda High PV perturbations is performed. Two criteria need to be met for a case to be composited. The case needs to be a homogeneous case as to not bias any one model, and the case needs to exhibit high DLMSF error. For 24 hr lead time, there are 149 homogeneous cases, with 65 of these cases exceeding 1.5 standard deviations of DLMSF error. The high error threshold was calculated by finding the DLMSF error standard deviation of each model, then taking the mean over all models. 1.5 standard deviations was chosen as a compromise between sufficiently large error and enough homogeneous cases to average for a meaningful composite. A forecast from all four models must surpass this individual threshold to be considered in the high error composite. The aforementioned steps to obtain a high error composite were performed individually for 24, 48, and 72 hr lead time. The composite at 24 hr lead time is shown in Figure 16. The contours show the average analysis balanced geopotential height field from the Bermuda High which can be thought of as the observed balanced mean field. The shading shows the average error in the balanced geopotential height field for the homogeneous cases, or deviations from the observed balanced mean field. Storm locations for the 65 high error cases are plotted as well. From Figure 16, all models show a weak bias in the balanced geopotential height field in the western periphery of the Bermuda High, which indicates that the strength of the Bermuda

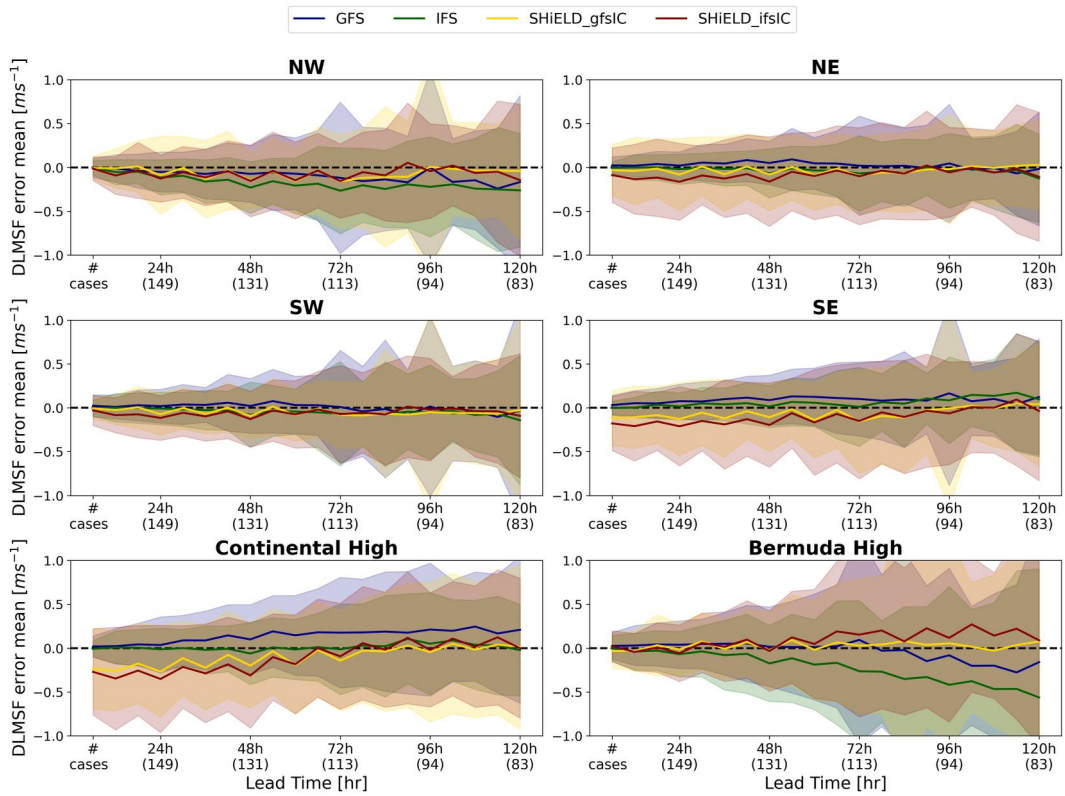

Figure 14. DLMSF error mean (bias) plotted as a function of system and forecast time. The bias is plotted every 6 hr out to 5 days. The same homogeneous cases as in Figure 13 are used to plot the bias. The colored shading indicates the $\pm 1\sigma$ range of DLMSF errors that went into calculating the bias point for each model.

High is generally underrepresented at 24 hr lead time in these forecasts. Colored markers illustrate hurricane locations following a recurving hurricane track pattern and are consistent with the track pattern associated with a weakened Bermuda High. While all models show a weak bias in the balanced geopotential height field, the spatial pattern varies slightly between the models. The IFS shows the lower balanced geopotential height field confined mainly to the western periphery of the Bermuda High, while in the GFS lower balanced geopotential height field also extend to parts of the continental United States and the Eastern Pacific. The spatial pattern of the balanced geopotential height field in the SHiELD gfsIC is similar to the GFS, but with lower balanced geopotential heights. The SHiELD ifsIC showed lower balanced geopotential heights in the western periphery of the Bermuda High similar to the IFS. The weak bias was even stronger with the spatial pattern extending further across the North Atlantic.

Figure 17 shows the composite at 48 hr lead time. Similar spatial patterns in the balanced geopotential height error are observed at the 48 hr lead time composite as in the 24 hr lead time composite. Moreover, there is noticeably larger error in all of the models which is expected given that errors tend to accumulate and grow with time. However, the SHiELD ifsIC at 48 hr lead time shows lower balanced geopotential height error in the central region of the Bermuda High and near-zero error in the western periphery of the Bermuda High. The SHiELD ifsIC also shows larger errors in the mid-latitudes. Figure 18 shows the composite at 72 hr lead time. Larger errors are observed at 72 hr lead time compared to the earlier forecast times. Notably, the magnitudes of the errors are higher, but the spatial patterns of the errors are different. The higher errors at 72 hr lead time suggest that the differences in the models at longer lead times are less related to the initial conditions, and that the model dynamics and/or physics are starting to play a larger role in producing errors.

Chen et al. (2019) calculated the track error in forecasts of Hurricane Irma and Hurricane Maria in the GFS, IFS, SHiELD gfsIC, and SHiELD ifsIC. The average total track error over all the forecasts for each storm was decomposed into an along-track error which is associated with the TC translation speed and a cross-track error

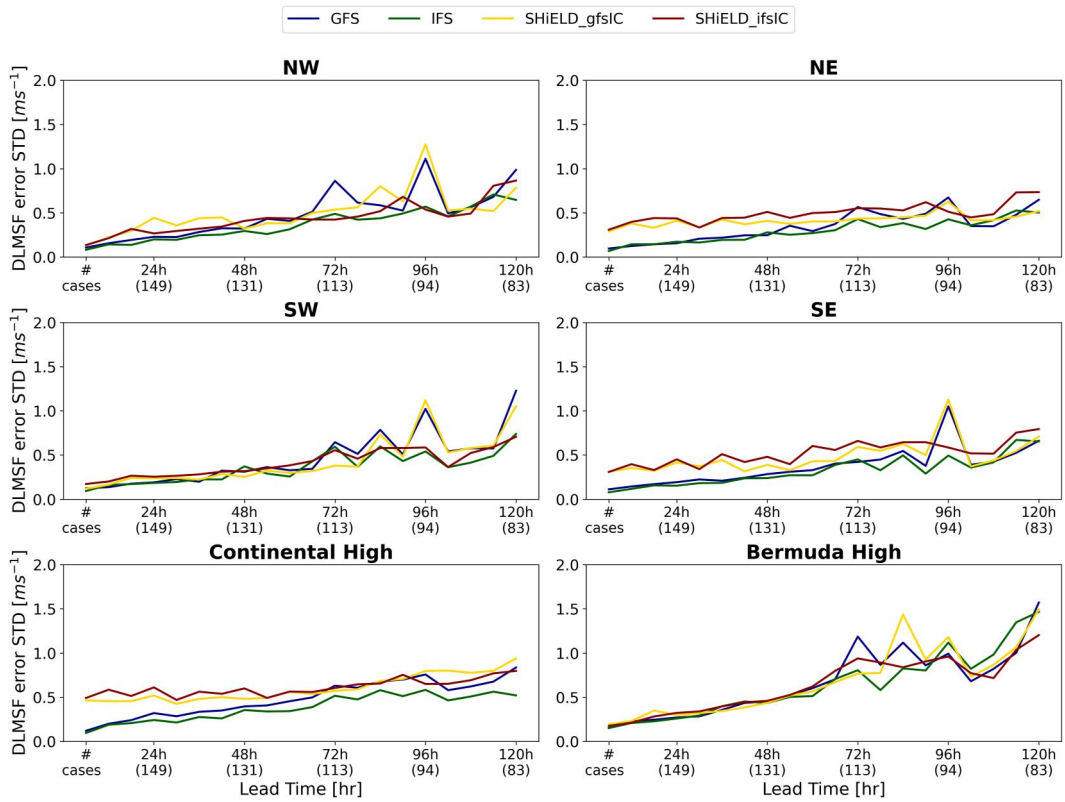

Figure 15. DLMSF error standard deviation plotted as a function of system and forecast time. The standard deviation is plotted every 6 hr out to 5 days. The standard deviation is calculated using the same homogeneous cases as in Figure 13.

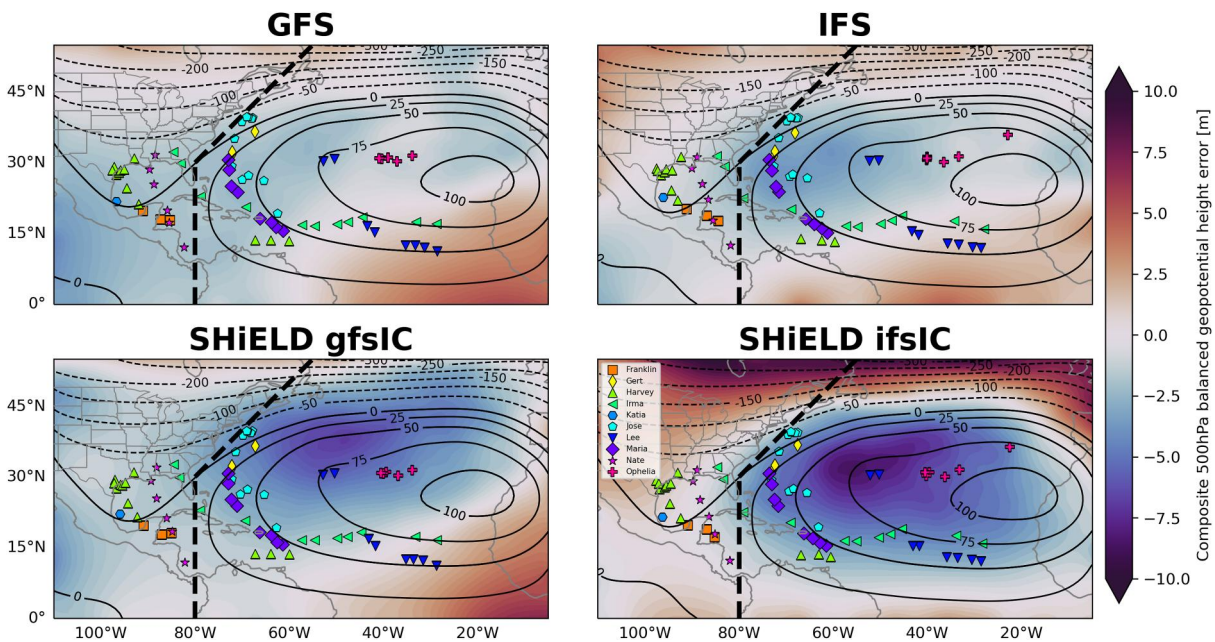

Figure 16. Composite analysis of 500 hPa balanced geopotential height analysis field (contours, m) and balanced geopotential height bias (shading, m) for 65 homogeneous cases exceeding 1.5 standard deviations of DLMSF error for the Bermuda High at 24 hr lead time. The 65 cases are plotted with markers denoting the storm's positions (markers).

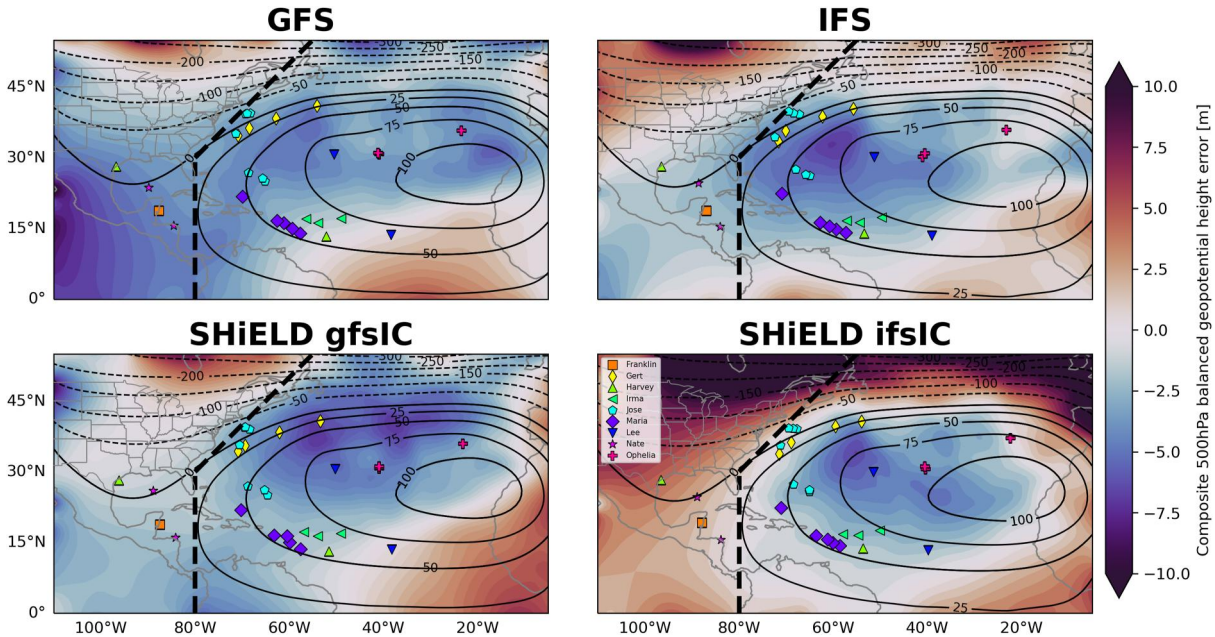

Figure 17. As in Figure 16, but for 48 hr lead time. Twenty-nine homogeneous cases exceeding 1.5 standard deviations of DLMSF error for the Bermuda High are displayed.

component which is related to the TC directional change. The IFS along-track error was consistently a significant fraction of the total track error, whereas the along-track and cross-track errors were more evenly split in the SHiELD ifsIC for both storms. In conclusion, they found that the along-track error showed a strong negative bias at all forecast times, meaning that the translation speed was consistently weaker in the IFS forecast compared to the verified track. The findings in this paper support the conclusion of Chen et al. (2019) and show that the slow translation speed bias in the IFS may be caused in part by errors in the forecast of the Bermuda High in this subset of forecasts.

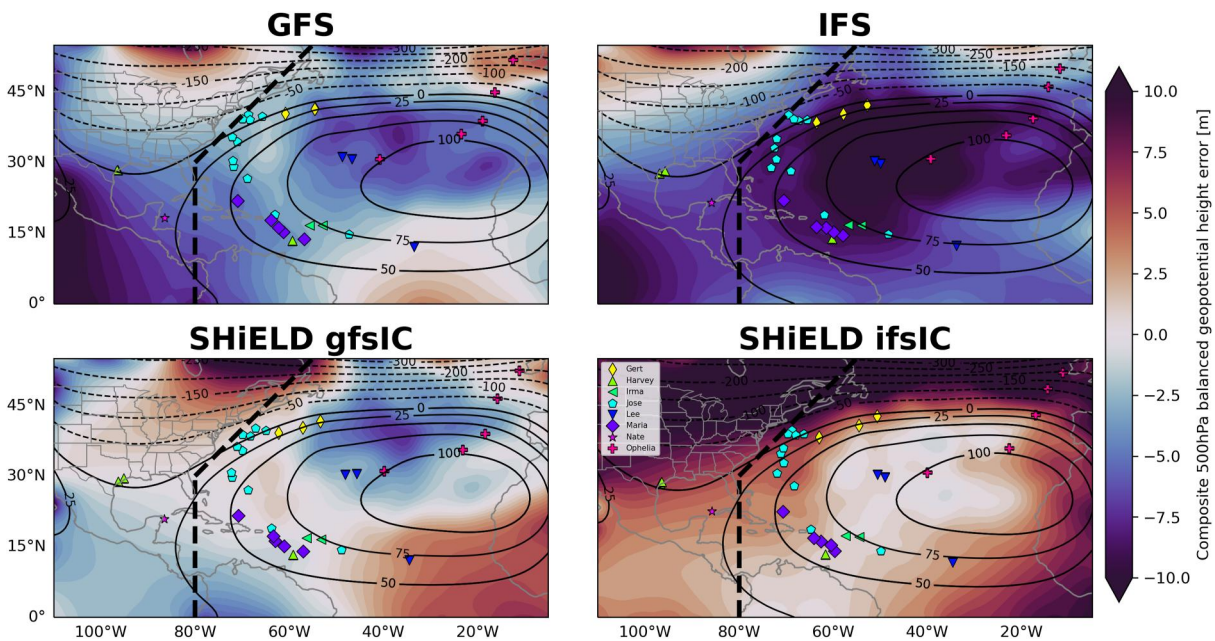

Figure 18. As in Figure 16, but for 72 hr lead time. Thirty-three homogeneous cases exceeding 1.5 standard deviations of DLMSF error for the Bermuda High are displayed.

5. Conclusions and Discussion

We utilize the piecewise potential vorticity (PV) inversion diagnostic technique to identify the sources of errors in TC track forecasts. Specifically, the contributions from individual large-scale pressure systems to TC steering flow and TC track errors are quantified. The PV framework in Wu et al. (2003) was adopted and altered to perform piecewise PV inversion systematically on the peak months of the 2017 Atlantic hurricane season. The perturbation PV was isolated by removing the TC from the total PV field, therefore the influence of the large-scale environment on TC movement was quantified. The perturbation PV field was partitioned in six regions geographically by sign: the Bermuda High, Continental High, and four quadrants based on the TC location. The nonlinear balanced flow was inverted from each perturbation PV piece individually to break down the environmental wind into its core components. The Deep Layer Mean Steering Flow (DLMSF) metric was calculated to quantify the contributions from these individual pressure systems on TC movement. DLMSF errors were computed by comparing the forecast to analysis fields. Lead-lag correlation analyses between DLMSF errors and track errors showed a maximum correlation from −24 to 0 hr, suggesting that track errors were associated with errors in the TC DLMSF up to 24 hr before. The correlations were moderate, suggesting that DLMSF is a sufficient proxy for TC movement and that DLMSF errors are generally representative of TC movement errors in model forecasts.

An in-depth analysis of steering flow from forecasts of Hurricanes Harvey, Irma, and Maria was performed using our piecewise PV inversion diagnostic. Harvey's weak steering environment was caused by two anticyclones contributing opposing flow, which ultimately led to its stall. The GFS forecast exhibited a westward track error attributed to zonal errors in the Continental High which favored westward movement. Irma displayed a northward track error in the SHiELD gfsIC forecast, placing its landfalling location near Miami as opposed to southwest Florida. The northward track error was mainly caused by DLMSF errors from the Continental High and Bermuda High, which imparted opposing meridional steering, and an overall net southerly steering on Irma. Maria's track in the IFS forecast was characterized by a southward track error, which resulted in the forecasted hurricane not actually making landfall in Puerto Rico. The DLMSF diagnosis showed opposing meridional steering on Maria from various systems, but a net larger northerly steering error resulted in Maria's southward track error. Large errors from Hurricane Jose to the northwest of Maria likely also played a part in the steering. Yet, the impact of error in Jose on Maria's movement was ambiguous because Jose existed on the boundary between the selected PV regions, and the DLMSF error from Jose was therefore split between these systems.

A basin-scale analysis of thirteen storms was performed to ascertain model biases. The DLMSF Mean Absolute Error (MAE) was calculated for all partitioned PV systems and showed that the model forecast errors of the Bermuda High contributed the largest TC movement errors on average, followed by the Continental High. The model forecast errors in the four quadrants contributed relatively small TC movement errors. The DLMSF error bias attributed to the Bermuda High was the largest, but varied in each model. To understand why the Bermuda High produced the most DLMSF MAE, homogeneous cases with DLMSF error more than 1.5 standard deviation associated with the Bermuda High were composited at 24-, 48-, and 72-hr forecast lead times. Compared to the analyses, the composites for all models showed lower balanced geopotential heights near and around the western side of the Bermuda High at 24 hr lead times and longer, indicating that all models consistently underrepresented the strength of the Bermuda High in erroneous forecasts. While all models showed lower geopotential heights on average, slight spatial differences in the composite field were apparent. The IFS showed lower balanced geopotential heights confined to the western periphery of the Bermuda High, while in the GFS lower balanced geopotential height also extended to parts of the continental United States and the Eastern Pacific. The spatial pattern of the balanced geopotential height field in the SHiELD gfsIC was similar to the GFS, but with lower balanced geopotential heights. The SHiELD ifsIC showed lower balanced geopotential heights in the western periphery of the Bermuda High similar to the IFS, but the weak bias was even stronger, and the spatial pattern extended further across the North Atlantic. The storm locations for the high error cases also showed a tendency for TCs to recurve, consistent with observed storm movement in the presence of a weaker Bermuda High in the North Atlantic. The results here support the findings in Chen et al. (2019) where the slow translation speed bias in the IFS is attributed to an underrepresented Bermuda High in the IFS forecasts providing weaker anticyclonic steering on TCs.

The 2017 North Atlantic hurricane season provides more than 100 TC forecast cases, which provides an essential proof of concept for using this diagnostic technique for assessing model biases in track error. While it would be

ideal to have multiple years to include a larger sample size to increase the significance of our study, the predicament of model consistency across multiple seasons becomes an issue. Also, it may be of interest to include other ocean basins to identify model biases of TC movement in different regions. We acknowledge that selecting only a subset of forecasts does indeed reveal a specific bias compared to averaging over a much larger sample of forecasts. However, we believe that the fact there are systematic patterns in the height errors does suggest that we can gain some insight into the representation of the Bermuda High from this subset rather than the expected near-zero long-term average over many forecasts with and without TCs present.

The systematic piecewise PV inversion tool we developed can efficiently analyze the track errors for multiple storms. However, it has some limitations. The choice of PV perturbation partitioning scheme cannot be entirely dependent on the atmospheric feature that is being analyzed. As was shown in the case study of Hurricane Maria, errors in the representation of Hurricane Jose to Maria's northwest existed on the boundary separating the PV regions. Consequently, the DLMSF error from the misrepresentation of Hurricane Jose was not obvious to interpret. The inability to identify the influence on a TC from an individual PV system in time is a shortcoming of the current partitioning scheme. A more complex algorithm would be required to analyze the contribution from individual systems. Still, the current partitioning method provides a first step toward a general partitioning scheme.

Our analysis only considers the advection from large-scale systems to TC movement. However, in the real atmosphere, a TC's intensity may also contribute to its own movement through a two-way interaction with its environment, known as the Beta effect (Marks, 1992). The Shapiro decomposition removes the intensity aspect by removing the TC from the total PV field, and only the perturbation field is used to quantify the DLMSF. Any feedback between the intensity and movement is thus eliminated, and the advection of TCs by their environmental steering can be looked at as a one-way interaction. Future work may include the contribution to TC movement from the TC PV field itself.

The lead-lag correlation analysis in Section 2.3 showed moderate correlations between DLMSF error and track error, and thus the DLMSF from 925 to 300 hPa is a reasonable proxy for TC motion. However, the instantaneous height of the TC vortex could be important in determining which atmospheric layers contribute to its movement. For a shallow vortex in the real atmosphere, the steering levels above the height of the vortex should not have a large impact on TC motion. However, the DLMSF does not account for varying vortex height and may over-contribute steering from atmospheric layers when the vortex is weak, which may cause the DLMSF to not accurately predict TC motion at times. A more accurate representation of the steering flow may be calculated by incorporating vortex height into the steering flow calculation.

While several previous studies have used the piecewise PV inversion diagnostic tool on single cases of TC movement, this is the first study to use piecewise PV inversion systematically over the peak months of a whole hurricane season to investigate model biases. Research presented here has shown that the piecewise PV inversion technique is a practical diagnostic tool in understanding TC track forecast errors in models. Improvement of our systematic piecewise PV inversion diagnostic toward a more comprehensive tool will provide further concrete evidence of biases present in model forecasts. The identification of model biases is the first step to rectifying them and improving TC track forecast errors.

Data Availability Statement

We adopted the piecewise PV inversion technique utilized in Wu et al. (2003). Post-processed data files, plotting scripts, and our modified piecewise PV inversion code can be found at: http://doi.org/10.6084/m9.figshare.24076680.

Acknowledgments
We express our gratitude to two anonymous reviewers whose comments helped improve the interpretation of the results and the overall paper. Additionally, our appreciation extends to Professor Chun-Chieh Wu's group at the National Taiwan University for providing the original piecewise PV inversion code. We also thank Dr. Treng-Shi Huang from the Central Weather Administration in Taiwan for their valuable discussion of the piecewise PV inversion methodology. This study was supported by funding from the Cooperative Institute for Modeling the Earth System, the G. Unger Vetlesen Foundation, and the Office of Naval Research Award N000142012069.

References

Baxter, M. A., Schumacher, P. N., & Boustead, J. M. (2011). The use of potential vorticity inversion to evaluate the effect of precipitation on downstream mesoscale processes. *Quarterly Journal of the Royal Meteorological Society*, *137*(654), 179–198. https://doi.org/10.1002/qj.730

Blake, E. S., & Zelinsky, D. A. (2018). National Hurricane Center tropical cyclone report Hurricane Harvey (Vol. *77*). Retrieved from https://www.nhc.noaa.gov/data/tcr/AL092017_Harvey.pdf

Cangialosi, J. P., Latto, A. S., & Berg, R. (2021). National Hurricane Center tropical cyclone report Hurricane Irma (Vol. *111*). Retrieved from https://www.nhc.noaa.gov/data/tcr/AL112017_Irma.pdf

Chan, J. C. L., & Gray, W. M. (1982). Tropical cyclone movement and surrounding flow relationships. *Monthly Weather Review*, *110*(10), 1354–1374. https://doi.org/10.1175/1520-0493(1982)110<1354:TCMASF>2.0.CO;2

Charney, J. (1955). The use of the primitive equations of motion in numerical prediction. *Tellus*, *7*(1), 22–26. https://doi.org/10.3402/tellusa.v7i1.8772

Chen, G. T.-J., Wang, C.-C., & Liu, S. C.-S. (2003). Potential vorticity diagnostics of a Mei-yu front case. *Monthly Weather Review*, *131*(11), 2680–2696. https://doi.org/10.1175/1520-0493(2003)131<2680:PVDOAM>2.0.CO;2

Chen, J.-H., Chen, X., Lin, S.-J., Magnusson, L., Bender, M., Zhou, L., & Rees, S. (2018). Tropical cyclones in GFDL fvGFS—Impacts of dycore, physics and initial conditions. In *33rd Conf. on Hurricane and Tropical Meteorology*.

Chen, J.-H., Lin, S.-J., Linus, M., Bender, M., Chen, X., Zhou, L., et al. (2019). Advancements in hurricane prediction with NOAA's next-generation forecast system. *Geophysical Research Letters*, *46*(8), 4495–4501. https://doi.org/10.1029/2019gl082410

Davis, C. A. (1992a). Piecewise potential vorticity inversion. *Journal of the Atmospheric Sciences*, *49*(16), 1397–1411. https://doi.org/10.1175/1520-0469(1992)049<1397:PPVI>2.0.CO;2

Davis, C. A. (1992b). A potential-vorticity diagnosis of the importance of initial structure and condensational heating in observed extratropical cyclogenesis. *Monthly Weather Review*, *120*(11), 2409–2428. https://doi.org/10.1175/1520-0493(1992)120<2409:APVDOT>2.0.CO;2

Davis, C. A., & Emanuel, K. A. (1991). Potential vorticity diagnostics of cyclogenesis. *Monthly Weather Review*, *119*(8), 1929–1953. https://doi.org/10.1175/1520-0493(1991)119<1929:PVDOC>2.0.CO;2

DeHart, J. C., & Bell, M. M. (2020). A comparison of the polarimetric radar characteristics of heavy rainfall from Hurricanes Harvey (2017) and Florence (2018). *Journal of Geophysical Research: Atmospheres*, *125*(11), e2019JD032212. https://doi.org/10.1029/2019JD032212

ECMWF. (2016). *IFS documentation CY41R2 - Technical report*. ECMWF. https://doi.org/10.21957/9phrx9eu

Ertel, H. (1942). Ein neuer hydrofynamischer wirbelsatz. *Meteorologische Zeitschrift*, *59*, 277–282.

Hakim, G. J., Keyser, D., & Bosart, L. F. (1996). The Ohio Valley wave-merger cyclogenesis event of 25–26 January 1978. Part II: Diagnosis using quasigeostrophic potential vorticity inversion. *Monthly Weather Review*, *124*(10), 2176–2205. https://doi.org/10.1175/1520-0493(1996)124<2176:TOVWMC>2.0.CO;2

Harris, L., Zhou, L., Lin, S.-J., Chen, J.-H., Chen, X., Gao, K., et al. (2020). GFDL SHiELD: A unified system for weather-to-seasonal prediction. *Journal of Advances in Modeling Earth Systems*, *12*(10), e2020MS002223. https://doi.org/10.1029/2020ms002223

Hoskins, B. J., McIntyre, M. E., & Robertson, A. W. (1985). On the use and significance of isentropic potential vorticity maps. *Quarterly Journal of the Royal Meteorological Society*, *111*(470), 877–946. https://doi.org/10.1256/smsqj.47001

Huo, Z., Zhang, D.-L., & Gyakum, J. R. (1999). Interaction of potential vorticity anomalies in extratropical cyclogenesis. Part I: Static piecewise inversion. *Monthly Weather Review*, *127*(11), 2546–2562. https://doi.org/10.1175/1520-0493(1999)127<2546:IOPVAI>2.0.CO;2

Klotzbach, P. J., Collins, C. J. J. M., III, Collins, J. M., Bell, M. M., Blake, E. S., & Roache, D. (2018). The extremely active 2017 North Atlantic hurricane season. *Monthly Weather Review*, *146*(10), 3425–3443. https://doi.org/10.1175/MWR-D-18-0078.1

Korner, S. O., & Martin, J. E. (2000). Piecewise frontogenesis from a potential vorticity perspective: Methodology and a case study. *Monthly Weather Review*, *128*(5), 1266–1288. https://doi.org/10.1175/1520-0493(2000)128<1266:PFFAPV>2.0.CO;2

Lackmann, G. M. (2002). Cold-frontal potential vorticity maxima, the low-level jet, and moisture transport in extratropical cyclones. *Monthly Weather Review*, *130*(1), 59–74. https://doi.org/10.1175/1520-0493(2002)130<0059:CFPVMT>2.0.CO;2

Landsea, C. W., & Cangialosi, J. P. (2018). Have we reached the limits of predictability for tropical cyclone track forecasting? *Bulletin of the American Meteorological Society*, *99*(11), 2237–2243. https://doi.org/10.1175/BAMS-D-17-0136.1

Marks, D. G. (1992). The beta and advection model for hurricane track forecasting. Retrieved from https://repository.library.noaa.gov/view/noaa/7184

Miller, R. J., Schrader, A. J., Sampson, C. R., & Tsui, T. L. (1990). The automated tropical cyclone forecasting system (ATCF). *Weather and Forecasting*, *5*(4), 653–660. https://doi.org/10.1175/1520-0434(1990)005<0653:tatcfs>2.0.co;2

Morgan, M. C. (1999). Using piecewise potential vorticity inversion to diagnose frontogenesis. Part I: A partitioning of the Q vector applied to diagnosing surface frontogenesis and vertical motion. *Monthly Weather Review*, *127*(12), 2796–2821. https://doi.org/10.1175/1520-0493(1999)127<2796:UPPVIT>2.0.CO;2

NCEP. (2016). List of GFS implementations - 3526 – GFS v14.0.0. Retrieved from https://www.emc.ncep.noaa.gov/emc/pages/numerical_forecast_systems/gfs/implementations.php

Pasch, R. J., Penny, A. B., & Berg, R. (2023). National Hurricane Center tropical cyclone report Hurricane Maria (Vol. *48*). Retrieved from https://www.nhc.noaa.gov/data/tcr/AL152017_Maria.pdf

Rappaport, E. N., Franklin, J. L., Avila, L. A., Baig, S. R., Beven, J. L., Blake, E. S., et al. (2009). Advances and challenges at the National Hurricane Center. *Weather and Forecasting*, *24*(2), 395–419. https://doi.org/10.1175/2008WAF2222128.1

Rossby, C.-G. (1940). Planetary flow pattern in the atmosphere. *Quarterly Journal of the Royal Meteorological Society*, *66*(S1), 68–87. https://doi.org/10.1002/j.1477-870x.1940.tb00130.x

Sampson, C. R., & Schrader, A. J. (2000). The automated tropical cyclone forecasting system (version 3.2). *Bulletin of the American Meteorological Society*, *81*(6), 1231–1240. https://doi.org/10.1175/1520-0477(2000)081<1231:tatcfs>2.3.co;2

Schlemmer, L., Martius, O., Sprenger, M., Schwierz, C., & Twitchett, A. (2010). Disentangling the forcing mechanisms of a heavy precipitation event along the alpine south side using potential vorticity inversion. *Monthly Weather Review*, *138*(6), 2336–2353. https://doi.org/10.1175/2009MWR3202.1

Shapiro, L. J. (1996). The motion of Hurricane Gloria: A potential vorticity diagnosis. *Monthly Weather Review*, *124*(11), 2497–2508. https://doi.org/10.1175/1520-0493(1996)124<2497:TMOHGA>2.0.CO;2

Simpson, R. H. (1974). The hurricane disaster—Potential scale. *Weatherwise*, *27*(4), 169–186. https://doi.org/10.1080/00431672.1974.9931702

Tamarin, T., & Kaspi, Y. (2016). The poleward motion of extratropical cyclones from a potential vorticity tendency analysis. *Journal of the Atmospheric Sciences*, *73*(4), 1687–1707. https://doi.org/10.1175/JAS-D-15-0168.1

Wang, C.-C., Lin, K.-Y., Davis, C. A., Huang, S.-Y., Liu, S. C.-S., Tsuboki, K., & Jou, B. J.-D. (2020). A modeling study on the impacts of Typhoon Morakot's (2009) vortex structure on rainfall in Taiwan using piecewise potential vorticity inversion. *Journal of the Meteorological Society of Japan. Ser. II*, *98*(4), 707–733. https://doi.org/10.2151/jmsj.2020-036

Winters, A. C., Keyser, D., & Bosart, L. F. (2020). Composite vertical-motion patterns near North American Polar–Subtropical Jet superposition events. *Monthly Weather Review*, *148*(11), 4565–4585. https://doi.org/10.1175/MWR-D-20-0140.1

Winters, A. C., & Martin, J. E. (2017). Diagnosis of a North American Polar–Subtropical Jet superposition employing piecewise potential vorticity inversion. *Monthly Weather Review*, *145*(5), 1853–1873. https://doi.org/10.1175/MWR-D-16-0262.1

Wu, C.-C., Chen, S.-G., Yang, C.-C., Lin, P.-H., & Aberson, S. D. (2012). Potential vorticity diagnosis of the factors affecting the track of Typhoon Sinlaku (2008) and the impact from dropwindsonde data during T-PARC. *Monthly Weather Review*, *140*(8), 2670–2688. https://doi.org/10.1175/MWR-D-11-00229.1

Wu, C.-C., & Emanuel, K. A. (1995a). Potential vorticity diagnostics of hurricane movement. Part I: A case study of Hurricane Bob (1991). *Monthly Weather Review*, *123*(1), 69–92. https://doi.org/10.1175/1520-0493(1995)123<0069:PVDOHM>2.0.CO;2

Wu, C.-C., & Emanuel, K. A. (1995b). Potential vorticity diagnostics of hurricane movement. Part II: Tropical Storm Ana (1991) and Hurricane Andrew (1992). *Monthly Weather Review*, *123*(1), 93–109. https://doi.org/10.1175/1520-0493(1995)123<0093:PVDOHM>2.0.CO;2

Wu, C.-C., Huang, T.-S., & Chou, K.-H. (2004). Potential vorticity diagnosis of the key factors affecting the motion of Typhoon Sinlaku (2002). *Monthly Weather Review*, *132*(8), 2084–2093. https://doi.org/10.1175/1520-0493(2004)132<2084:PVDOTK>2.0.CO;2

Wu, C.-C., Huang, T.-S., Huang, W.-P., & Chou, K.-H. (2003). A new look at the binary interaction: Potential vorticity diagnosis of the unusual southward movement of Tropical Storm Bopha (2000) and its interaction with Supertyphoon Saomai (2000) [Software]. Monthly Weather Review, *131*(7), 1289–1300. https://doi.org/10.1175/1520-0493(2003)131<1289:ANLATB>2.0.CO;2

Wu, C.-C., & Kurihara, Y. (1996). A numerical study of the feedback mechanisms of hurricane–environment interaction on hurricane movement from the potential vorticity perspective. *Journal of the Atmospheric Sciences*, *53*(15), 2264–2282. https://doi.org/10.1175/1520-0469(1996)053<2264:ANSOTF>2.0.CO;2