2. ABSTRACT

Application of Innovation Statistics to Diagnose Biases in the HAFS system
Professor Ryan Torn, University at Albany, SUNY
Total Cost: $292,483, 1 September 2020 - 31 August 2022
Collaborators: NCEP/EMC, AOML/HRD

Numerical weather prediction models are complex sets of equations designed to represent or parameterize numerous processes that dictate the evolution of the atmosphere. In the Hurricane Analysis and Forecasting System (HAFS), these equations are used to predict the evolution of tropical cyclone (TC) position, intensity, and eventually hazards, such as precipitation and wind. These models frequently contain regional or process-based biases, which in turn can yield biases in TC-related predictions. As a consequence, it is important to identify and alleviate the source of the biases, which can be difficult due to the complex interplay of different model components. This proposal, which primarily addresses Priority 4: "Develop improvements for tropical cyclone predictions, especially reducing hurricane intensity errors and improving the predictions for rapid changes in intensity”, aims to use statistics of the difference between short-term model forecasts and observations (i.e., innovation statistics) to identify and alleviate the source of model biases. In this work, we will develop a software framework that compares sets of short-term HAFS forecasts against a variety of observations that would allow the user to identify biases related to the large-scale environment and TC vortex. Large-scale biases will be diagnosed by comparing HAFS forecasts against regular observations, such as rawinsondes or satellite winds, in an earth-relative framework, which can permit the identification of regional and/or phenomenological biases that can impact TC track or the environment. For TC intensity, we plan to compare HAFS forecasts against dropwindsondes, radar data, and other TC-specific observations. These comparisons will be done in a storm and shear-relative framework, which will facilitate computing statistics over many storms. Averaging the bias over many cases will identify where the model issues lie and suggest where improvements should be made and help make better use of observations within a data assimilation system. The results of these calculations will be communicated to the HAFS model developers, which will help them to identify priority areas for HAFS development. Finally, the techniques developed here can be extended to identifying biases in other UFS applications; therefore, the usefulness of this research can extend beyond HAFS. At the end of the project, we plan to make the software package available to the broader HAFS and UFS communities.
3. RESULTS FROM PRIOR RESEARCH

PI Torn has received funding during the 2012, 2014, 2016, and 2018 HFIP Announcement of Opportunity (AO) and has been an active participant in HFIP research since 2008. This research has generally focused on applications of tropical cyclone (TC) ensemble forecasting, understanding the role of large-scale errors on TC track and intensity forecasts, determining how initial condition errors impact TC intensity forecasts, and developing new methods of treating model error in ensemble forecasts.

The first HFIP award ("Evaluating Hurricane Intensity Predictability using the Advanced Hurricane WRF", NA12NWS4680003, 1 January 2012-31 December 2013, $150,000) consisted of tasks related to understanding the growth of intensity forecast uncertainty, diagnosing the source of Hurricane Sandy’s track uncertainty and developing an error diagnostic framework for evaluating conditional biases. The first task evaluated how uncertainty associated with atmospheric and oceanic initial conditions and model uncertainty limit the predictability of TC intensity via a series of ensemble experiments consisting of cases from four years. The results (described in Torn (2016)) suggested that atmosphere uncertainty lead to the largest intensity uncertainty during the first 72 h; however, ocean uncertainty alone lead to comparable intensity uncertainty by 96 h. By contrast, intensity uncertainty due to exchange coefficient variability leveled off in the first 48 h. Beyond that work, we evaluated the physical processes responsible for the large divergence in GEFS Hurricane Sandy position forecasts (collaboration with Drs. Tom Hamill and Jeff Whitaker, NOAA/ESRL Torn et al. 2015). The ensemble members with an eastern solution were characterized by less upper-tropospheric outflow due to convection 200-400 km to the north of Sandy, which in turn lead to a less-amplified ridge on the poleward side of Sandy, imparting a westerly wind on Sandy that gradually nudged the TC away from North America. This result suggests that forecasts could have been improved given better observational sampling of the upper-tropospheric outflow to the north of poleward-moving TCs. Finally, the project lead to the development of a verification package for assessing conditional TC track and intensity biases.

PI Torn’s second HFIP AO Award ("Evaluating Hurricane Intensity Predictability using the Advanced Hurricane WRF", NA14NWS4680027, 1 August 2014-31 July 2016, $337,385, PI: Ryan Torn) developed new applications of HWRF ensemble forecasts and novel methods for identifying biases. PI Torn has acted as the co-lead (with Dr. Mark DeMaria, NHC) of a HFIP Tiger Team devoted to providing NHC forecasters with additional guidance using output from the various HFIP TC ensemble forecasting systems. This activity lead to the creation of an ensemble-based rapid intensification probability, which is similar to the SHIPS-based product, but is computed from the ensemble distribution. As part of this project, PI Torn developed a method to bias-correct the HWRF ensemble intensity changes, software to compute the probabilities and output them in an ATCF-friendly format, and coordinated the data dissemination among the varied groups. During Hurricanes Irma and Nate (2017), PI Torn provided real-time guidance on the sensitivity of TC track forecasts to the steering flow to Mike Brennan and Eric Blake (NHC forecasters). This information was subsequently used as input for G-IV flight patterns for these storms. This resulted in a successful Joint Hurricane Testbed proposal ("Transitioning Ensemble-based TC Track and Intensity Sensitivity to Operations", NA19OAR4590129, 1 July 2019-30 June 2021, $149,880, PI: Ryan Torn). This award is facilitating the transition of PI Torn’s software into the python framework, while PI Torn continues to provide sensitivity support in real-time. The other major activity associated with this award has been to apply the large library of retrospective HWRF...
forecasts to diagnose the common processes responsible for large short-term forecast errors. In particular, we compared HWRF forecasts with large short-term intensity errors against analog forecasts with small short-term intensity errors. The results from this work indicate that the large error cases have less spatially-extensive and shallower convection, and lower boundary layer equivalent potential temperature relative to their analog low error counterparts (Halperin and Torn 2018).

In the third HFIP AO Award (“Evaluating Methods of Parameterizing Model Error in the HWRF Ensemble Prediction System”, NA16NWS4680025, 1 September 2016-31 August 2018 [in no cost extension], $328,574, PI: Ryan Torn) PI Torn’s group developed new methods of incorporating SST and model uncertainty into the HWRF ensemble. The former is accomplished by creating an ensemble of SST initial conditions by randomly selecting an SST state from the past, determining its deviation from a long-term climatology and adding the resulting SST perturbation to the control SST analysis. This method resulted in a 15% increase in intensity standard deviation relative to the control over four different TCs. The method was subsequently incorporated into the HWRF repository. The second aspect of this project is implementation of the Stochastic Perturbed Parameterization Tendency (SPPT) into the HWRF framework, which will be incorporated into the HWRF trunk soon. Our preliminary results suggest that applying SPPT to the PBL parameterization results in a 5-10% increase in the intensity standard deviation over four different storms. We are currently preparing a manuscript that describes the results of these experiments.

The final HFIP AO Award (“Evaluating Initial Condition Perturbation Methods in the HWRF Ensemble Prediction System”, NA18NWS4680060, 1 September 2018-31 August 2020, $292,483, PI: Ryan Torn), PI Torn’s group is evaluating the probabilistic hazard forecasts that can be derived from the HWRF ensemble system and comparing it against what can be derived from the GFS and ECMWF ensemble systems and the Monte Carlo Wind model. Preliminary analysis from 2017 and 2018 periods suggest that the ECMWF ensemble produces the most accurate probabilities, with GFS and HWRF worse. Ongoing work includes application of machine learning to improve probabilistic forecasts and continued interaction with NHC forecasters.

Outside of HFIP, PI Torn was a co-investigator on a NOAA award to the University of Miami to use the NASA Global Hawk to determine optimal sampling strategies for improving TC track and intensity forecasts (“Using NOAA UAS Assets and OSSE/DA Capabilities to Improve Sampling Strategies and Numerical Prediction of Tropical Cyclone Track, Intensity, and Structure”, NA14OAR4830172, 1 August 2014-31 July 2017, PI: Jason Dunion, $238,697 subaward to the University at Albany). During the Sensing Hazards with Operational Unmanned Technology (SHOUT) field project, PI Torn collaborated with EMC Scientist Dr. Zhan Zhang to generate a once-daily 80-member HWRF ensemble, which was used as input into the ensemble-based sensitivity algorithm to determine the optimal location(s) for deploying dropsondes that would reduce the uncertainty in TC position and intensity forecasts. These results were shared with SHOUT mission scientists prior to all flights both within the daily planning meeting and in a written report (available from: http://catalog.eol.ucar.edu/tci/150363/files). During the 2016 deployment, PI Torn carried out the calculations using output from both the HWRF and ECMWF ensemble prediction system. Both GFS and HWRF forecasts exhibit lower track and intensity errors when the SHOUT dropsondes are assimilated versus the forecasts where they are not (Jason Sippel, personal communication).
4. PROJECT DESCRIPTION

Introduction

Advances in scientific understanding have yielded a steady improvement in forecasts of tropical cyclone (TC) track; however, progress toward improving TC intensity forecasts has not been as great, particularly at lead times less than 3 d (e.g., Rogers et al. 2006; Rappaport et al. 2009). One of the potential reasons for the lack of significant improvement in TC intensity is that TC intensity is likely dictated by a combination of physical processes that occur on a number of space and time scales. These factors include large-scale dynamical factors, such as vertical wind shear, (e.g., DeMaria and Kaplan 1999; Kaplan and DeMaria 2003), vortex-scale asymmetries (e.g., Nolan and Grasso 2003; Rogers et al. 2013; Rios-Berrios and Torn 2017), as well as deep convective bursts near the TC core (e.g., Hendricks et al. 2010; Molinari and Volland 2010). The pioneering work of Lorenz (1969) suggests that the predictability timescale for a chaotic system like the atmosphere is proportional to the length scale of the feature being forecast. Given that the primary circulation of the TC is maintained by diabatic heating associated with convection, one might expect the predictability of TC intensity should be hours; however, this convection is occurring within a vortex that has a longer predictability timescale, which can act to constrain and organize the convection. Recent work suggests that the internal dynamics of mature TCs, which is often associated with moist convection, limits the predictability of TC intensity to 2 days or less (e.g., Hakim 2013; Brown and Hakim 2013). In contrast to this, Judt et al. (2015) found that high azimuthal wavenumber noise, which is essentially equivalent to adding uncertainty to convective elements inside the TC, had a relatively minor impact on the intensity of Hurricane Earl (2010). Instead, they found that only perturbations to azimuthal wavenumber 0 and 1 were able to impact Earl’s forecast intensity. In addition to the axisymmetric vortex, large-scale environmental factors, such as sea-surface temperature, vertical wind shear strongly modulate TC intensity, yet evolve on even slower timescales, which could provide longer-range predictability (e.g., Emanuel et al. 2004; Wang and Wu 2004). Moreover, 3D idealized simulations suggest that TC intensity predictability may be modulated by environmental forcing, such that moderate vertical wind shear and higher sea-surface temperature environments are the least predictable (e.g., Zhang and Tao 2013; Tao and Zhang 2014).

Given the complexity of predicting TC intensity, accurate forecasts of this metric will require employing numerical weather prediction models. In general, these systems produce predictions of the evolution of the atmosphere by applying the equations of motion for the atmosphere (conservation of momentum, mass, heat, and water) to the best estimate of the current state of the atmosphere (i.e., an analysis). For a variety of reasons, these models do not contain perfect representation of all of the processes present in the atmosphere. These reasons include the fact that the dynamical equations of motion are continuous differential equations, yet most models contain finite approximations to those terms. Furthermore, there are numerous processes that either occur on a smaller scale than what our models can represent (i.e., clouds, convection, turbulent mixing), or processes that are not accounted for by the main equations of motion (i.e., radiative transfer, land surface processes, ocean mixing). As a consequence, the net impact of these processes are represented through parameterization schemes. Model parameterizations use grid-resolved variables as input, such as the wind, temperature, and water vapor in a column, which in turn is converted into a forcing tendency for the momentum, heat, and moisture equations. The
relationships that map from model grid to tendencies is often based on statistical relationships, which are trained over a small number of case studies, or by optimizing forecast metrics over a limited set of cases. As a consequence, independent forecasts often exhibit errors and biases.

The complexity of these numerical models, including the number of grid points, and the interaction between parameterizations can make it difficult to identify the source of these biases and how to correct these. For example, the evolution of TC intensity within these models is a result of how well the model simulates the various processes that determine TC intensity, such as surface fluxes (e.g., Green and Zhang 2013), boundary layer mixing (e.g., Braun and Tao 2000; Kepert 2012), and cloud microphysics (e.g., Wang 2002; Zhu and Zhang 2006). As a consequence, it can be difficult to determine what aspect of a model is leading to biases in TC intensity. Furthermore, most of the development attention is paid toward the final metrics, with comparatively little attention toward the processes.

One method of assessing a model’s performance is to cycle observations using a data assimilation system. Within a data assimilation system, the model state is adjusted based on the difference between the observation value, and the model estimate of the observation, which is known as an innovation. The model estimate of the observation is computed using an observation operator, which maps from the model state vector into the observation, which can be as simple as bi-linear interpolation, or as complex as a radiative transfer operator that maps from a temperature column to a radiance that might be measured from a spaceborne satellite. Ideally, the innovation should be zero, which would mean that the model is replicating the observation perfectly. For any single observation, this quantity will be non-zero due to a number of factors, including observation errors (due to the instrument and how well the model represents the observation), random errors in the model (i.e., errors that are unique to a single forecast), and systematic biases. As a consequence, a single observation and/or time does not provide much information about the quality of the model itself. By contrast, the mean innovation over many observations should be zero; otherwise, it would indicate a systematic bias in the model. As a consequence, accumulating innovation statistics over many forecasts and/or locations can yield significant amounts of information about a model and where issues may be present. Furthermore, most data assimilation systems assume that the innovation values are unbiased. Some observations, such as satellite radiances are characterized by significant biases, which in turn has necessitated the development of sophisticated bias-correction methods to utilize this information, which has been instrumental in global NWP (e.g., Dee 2005).

Most NWP centers monitor innovations for individual forecasts as a means of assessing the quality of the data assimilation system and to identify and remove bad observations (e.g., Langland and Baker 2004; Cardinali 2009; Hotta et al. 2017). Moreover, other studies have used higher-moment statistics, such as the variance in the innovation statistics, to determine the appropriate error variances within a data assimilation system (e.g., Desrozières et al. 2005), or the appropriate inflation factor within an ensemble data assimilation system (e.g., Anderson 2009).

In addition to the above examples, there have been relatively few which have applied this approach to TCs. Torn and Davis (2012) demonstrated how innovation statistics could be used to identify biases within the Advanced Research version of WRF, which was cycled with observations over an active one month period during 2008. In this study, they computed mean innovations for the entire period as a function of horizontal and vertical location and used these output to determine the source of a westerly TC position bias in the forecast relative to the actual storm. In this study, the innovation statistics indicated that the model was consistently too warm in the subtropical ridge,
particularly in the western part of the Atlantic basin, which in turn was associated with stronger
easterly winds on the southern side of the subtropical ridge. The vertical distribution of innovation
statistics suggested that the largest biases were near the top of the boundary layer. The horizontal
and vertical distribution of biases suggested that the model was not simulating the effects of oceanic
shallow convective clouds, which redistribute heat and moisture through the trade wind inversion.
As a consequence, Torn and Davis (2012) replaced the Kain-Fritsch cumulus parameterization
with the Tiedtke scheme, which was characterized by smaller biases in the innovation statistics.
The latter contains a more appropriate treatment of shallow convective clouds in the tropics, while
the former was developed for southern Great Plains convection. In addition, forecasts initialized
with the Tiedtke scheme yielded track forecast errors that were up to 25% smaller than those with
the Kain-Fritsch scheme; therefore, it demonstrates a method by which the innovation statistics
applied to the model could be used to diagnose and correct deficiencies. Cavallo et al. (2016)
expanded upon this approach by combining innovation statistics with model tendencies to further
diagnose potential model biases and corrections.

The Torn and Davis (2012) focused on applying innovation statistics to the TC large-scale
environment; however, a similar approach has been used to diagnose and fix biases in the TC
core region. Zhang et al. (2015) compared short-term HWRF forecasts against dropwindsonde
data collected by aircraft that penetrated the TC core. Although this study did not specifically
use output from a data assimilation system, the resulting statistics are equivalent to evaluating
innovation statistics on dropwindsonde data. The resulting model differences suggested that the
HWRF turbulent mixing scheme was too vigorously mixing heat and momentum because the
mixing coefficient was too large. As a consequence, Zhang et al. (2015) reduced the mixing
coefficient by a constant factor $\alpha$, which in turn resulted in lower TC intensity errors, more skillful
RI forecasts Zhang et al. (2017), and exposed other biases in the model, resulting in better forecasts
of the TC structure, including the size, warm core, eyewall slope, etc. (Zhang et al. 2018). Despite
this, the TC-inner core data remains under-utilized to diagnose model biases beyond a few limited
case studies. This is critical, particularly from a data assimilation context because systematic biases
will reduce the impact of inner core data assimilation (e.g., Vukicevic et al. 2013; Tong et al. 2018).

Proposed work

Given the potential utility of innovation statistics in previous studies, the purpose of this proposal
is to develop the capability to routinely compute innovation statistics and stratify the results based
on geography and TC-relative scale. This in turn could help identify and address biases in various
components of the HAFS system, which in turn will ultimately improve TC track and intensity
forecasts. As a consequence, this proposal addresses Priority 4: “Develop improvements for
tropical cyclone predictions, especially reducing hurricane intensity errors and improving the
predictions for rapid changes in intensity”, with links to Priority 1b: “Develop process-based
diagnostics designed to improve forecasting across various scales”, and Priority 3a: “Test and
improve coupled UFS applications, identifying and addressing significant biases and coupling
shocks”. The proposed work will formalize collaborations between PI Torn and EMC scientist
Dr. Henry Winterbottom and HRD scientist Dr. Jason Sippel, who are closely involved in
the development of the HAFS data assimilation and modeling framework (see enclosed letter of
support). The technique has been demonstrated on a research-version model and data assimilation
system; therefore, the methodology is currently in either at RL-3 or RL-4. By the end of the project,
Table 1: List of observation types and domains over which each observation will be used to compute innovation statistics.

<table>
<thead>
<tr>
<th>Observation Source</th>
<th>Observation Types</th>
<th>Domains</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rawinsondes</td>
<td>wind, temperature, water vapor</td>
<td>earth-relative</td>
</tr>
<tr>
<td>Satellite-Derived Motion Vectors</td>
<td>wind</td>
<td>earth-relative</td>
</tr>
<tr>
<td>AIRS Retrievals</td>
<td>temperature, water vapor</td>
<td>earth-relative</td>
</tr>
<tr>
<td>Dropwindsondes</td>
<td>wind, temperature, water vapor</td>
<td>TC-relative</td>
</tr>
<tr>
<td>Doppler Velocities</td>
<td>wind</td>
<td>TC-relative</td>
</tr>
</tbody>
</table>

We expect to be at RL-6 since we will demonstrate the method in an end-to-end environment, with the possibility of demonstrating in an operational environment (RL-7).

The main tasks of this proposal involve two primarily development tasks: (i) creating software that can compare 6-h model forecasts against observations and (ii) establish a framework whereby the innovation statistics can be aggregated over many cases. Ideally, the innovation information, which involves comparing the observation value with the model estimate of the observation, should be carried out within the HAFS data assimilation system, with the option to output that information into a separate diagnostic file. A second python-based software framework would then ingest this information and subsequently compute and analyzes the innovation statistics in an appropriate coordinate framework(s). Data assimilation systems currently assimilate vast amounts of observations per cycle; however, not all of these observations will be useful for diagnosing model biases, due to either the distribution of observations, or whether the observations sample model fields. As a consequence, emphasis will be placed on observation types that are representative of model state variables, such as wind, temperature, and moisture, as well as observations that can characterize the 3D distribution of bias, including in both the horizontal and the vertical, on a regular basis. For example, rawinsondes sample wind, temperature, and moisture data at the same fixed locations, which makes this data valuable for this kind of study, while aircraft observations (i.e., ACARS) are not. Specifically, ACARS data can measure all three variables, but this data is concentrated at flight level and mainly within specific corridors of heavy aircraft traffic. Given the above restriction, the focus will be on computing and evaluating innovation statistics with respect to the observations listed in Table 1. It is worth pointing out that the methodology below could be expanded to any observation type, thus for brevity we are focusing on those below.

The most efficient way to carry out this work would be if the model estimates of the observation is computed within the HAFS data assimilation system; however, that component is not available at this point. Given the potential value of this approach, it is important to move this work forward before the HAFS data assimilation system is available. In the meantime, we will write software that manually computes the model estimate of each observation from gridded 6-h HAFS output. Most of these observations are model state variables; therefore, the operator that maps from the model variables to the observation value (which is necessary to compute the innovation statistics) is bilinear interpolation of variables from the model grid to the correct location of the observation. This will be necessary since most observations will not be located on model grid points. For ease, we will archive the observation type, location, value, and model estimate into a separate file, which will make it easier to use to compile statistics over a larger set of cases. In addition, we will work with
the HAFS data assimilation developers to make sure the data assimilation code contains a routine that outputs the observation and model estimate of the observation (i.e., innovation) into a separate file while the data assimilation system is running. Calculating the innovation information in the HAFS data assimilation system would allow future users to leverage the forward operators (i.e., the mapping from model state to observations), which reduces the long-term software maintenance. Furthermore, these quantities need to be computed within the HAFS data assimilation, thus it is more efficient to leverage this information. Furthermore, it will also allow for expanded use of innovation statistics for observations that have more complex forward operators (i.e., satellite radiances).

Once the software is available to compute innovation statistics, we will run it on the set of HAFS model forecasts that are made available to us by the developers. Ideally, we would have access to hundreds of forecast initialization times to carry out this analysis. The resulting innovation information files will be relatively small in size and easy to analyze. From the set of forecasts, we will aggregate the innovation statistics in two different frameworks to address biases within the large-scale environment around the TC and within the TC core using TC-relative framework. Here, the large-scale environmental biases will be diagnosed using an earth-relative framework. For observations that are deployed from the same place every day (i.e., rawinsondes), the innovation statistics for that location will be aggregated over time. By contrast, for observations that do not regularly sample the same location, (i.e., satellite-derived motion vectors), we will compile innovation statistics for all observations that fall within specified latitude/longitude boxes. The average innovation values can then be plotted on either on a 2D map on a single pressure level for a single variable, or in a column format, which is all vertical levels at a single horizontal location. The distribution of horizontal biases will then be used to identify aspects of the model state that are characterized by significant biases and hence what aspects of the model need to be addressed.

Fig. 1 shows an example of 700 hPa innovation statistics output from the AHW data assimilation system that was produced for separate work. While there appears to be minimal temperature bias over CONUS, the model is between 0.6 to 0.9 K colder than observations over much of the eastern Caribbean, though the same is not true over Africa, which is indicative of a regional bias (Fig. 1a). Furthermore, the zonal component of the wind is too westerly through the western Caribbean, which could indicate issues with the Caribbean low-level jet (Fig. 1b). By contrast, the winds in the eastern Atlantic Ocean are too southerly, which may indicate an issue with the subtropical ridge physics in that region, though we do not have good thermodynamic observations to confirm this (Fig. 1c). During this project, we would seek to generate similar style figures using HAFS output. As we generate these results, we will provide this output to the HAFS model and data assimilation developers and work with them to interpret the results and work with them to identify aspects of the model that might be giving rise to the biases. In the earth-relative framework, particular attention will be given to wind biases, which are likely to impact TC track error within the model, and moisture biases, which could yield TC intensity errors.

TCs are discrete events which tend to take place across an entire ocean basin. As a consequence, it is difficult to apply this method to evaluate model biases near the TC core in an earth-relative framework. Given that it is likely that HAFS will likely contain biases in processes that dictate TC intensity, it is of interest to develop innovation bias statistics in a TC-relative coordinate reference frame. Numerous previous studies have indicated that the direction of vertical wind shear produces significant asymmetries in TC structure and processes (e.g., Corbosiero and
Figure 1: (a) Mean difference between 6-h AHW 700 hPa temperature forecasts and rawinsonde observations during the 2009 hurricane season (units: K). Mean difference between 6-h AHW (b) zonal and (c) meridional wind and atmospheric motion vectors over the same period (units: m s\(^{-1}\)).

Molinari 2002; Reasor et al. 2013). As a consequence, we will compile innovation bias statistics in a storm-centered shear-relative framework, whereby each observation location is mapped into a polar coordinate framework, where the radius is the distance from the TC center at the time of the observation (determined from interpolation of best track data). Furthermore, the observation azimuth will be determined from the shear direction, which will be determined from SHIPS diagnostic files. Calculating biases in a storm-relative framework will elucidate biases that might exist in specific quadrants of the TC. In particular, we will investigate the upshear and right of shear quadrants, which are likely to impact TC intensity in a sheared storm since these are locations where surface fluxes are likely to impact downwind convection (e.g., Nguyen et al. 2019). Furthermore, we will pay particular attention to the boundary layer biases and wind field. The former has been shown to produce significant intensity errors in HWRF (Zhang et al. 2015), while the former is likely to impact the resiliency of the vortex to shear, which can lead to intensity biases. Furthermore, we will also assess whether there are intensity-specific biases by stratifying the innovation statistics for TCs whose intensity is below tropical storm strength (< 64 knots) and those who are at hurricane strength and above (>= 64 knots). Previous studies (e.g., Bhatia and Nolan 2013) have indicated that sheared tropical storms are often characterized by the largest errors; therefore, it is worthwhile to investigate whether there are intensity-specific biases in the
Plotting the biases in both horizontal and vertical sections will be effective in showing where the model is not consistent with observations; however, it does not necessary say how these biases might be related to each other. For example, balances between the mass and wind field might dictate that wind biases should be associated with corresponding temperature biases. While these relationships might be fairly straightforward to see in many instances, there are possible situations which might be harder to diagnose, particularly for vortex-scale TC observations that might be on different quadrants of the storm. For example, a boundary layer bias in the upshear quadrant could be associated with a wind bias in the downshear quadrant because the enhanced mixing in the upshear quadrant yields more vigorous convection, and in turn a stronger TC. The large number of cases that are necessary for this work opens up the possibility of using machine learning methods, such as random forests (e.g., Breiman 2001; Gagne et al. 2014), to diagnose relationships between the bias in one part of the storm vs. the bias from the other side of the storm. As a consequence, we will apply the random forest methodology to our innovation statistic dataset to diagnose the relationship between innovation biases. Here, the independent variables will be the position (i.e., lat., lon., for earth-relative, azimuth, radius for TC-relative), field (i.e., u, v, T), and forecast, while the dependent variable will be the average innovation bias. The goal of this calculation is not to produce a predictive model of what is the bias; rather, the purpose of this calculation is to analyze the coefficients that make up the relationships between variables and use it to diagnose the source of the bias.

Analyzing innovation statistics can also indirectly yield forecast improvements though data assimilation. As stated above, data assimilation systems assume unbiased models and observations; trying to assimilate observations with a biased model often leads to the observations temporarily correcting the model bias, which quickly returns within the forecast because the underlying source was not addressed. In turn, the promise of data assimilation, and in particular inner core data assimilation, is not realized (e.g., Vukicevic et al. 2013; Tong et al. 2018). Furthermore, this work will indicate locations and observation types that might be more problematic to assimilate into HAFS, which might suggest either withholding certain observations from the analysis system, or developing a bias correction scheme.

One of the appealing aspects of innovation statistics is that it is useful beyond the HAFS application; innovation statistics have potential utility in other UFS applications. Specifically, earth-relative innovation statistics can be useful for any UFS forecasting system that is run on a fixed domain, such as the GFS, or CAM applications. For example, the methodology could be extended to analyze 6-h forecast output from the GFS system and diagnose biases in different regions of the world, or in CAM applications to analyze regions of CONUS of larger PBL parameterization issues. Throughout the development process, we will work with HAFS and UFS developers to make sure that this diagnostic framework is generic and can be easily adopted by other UFS developers. In order to ensure this happens, we will make the software to compute the aggregated innovation statistics available through a public Github repository. We envision that innovation statistics could be part of the routine monitoring of a forecast system. Most of the calculations being proposed here are fairly straightforward; however, we will investigate whether it would make more sense to integrate this capability into the METplus framework. As of now, it does not appear this framework has the capability to compile observations in a TC and shear-relative framework, which would suggest using our own package, though we can be flexible.
Project Timeline

Year 1:
- Develop innovation statistics from HAFS forecast files
- Develop initial earth-relative innovation statistics
- Begin developing TC-relative innovation statistics capability

Year 2:
- Adapt innovation statistics output to work within HAFS data assimilation system
- Compute final earth-relative, and TC-relative innovation statistics
- Apply random forest methodology to innovation statistics
- Make code available to others through Github repository

5. DATA SHARING PLAN

This project will primarily employ pre-generated HAFS forecasts, which we will download from NOAA storage systems. The innovation statistics will be compiled from smaller text files that contain all of the necessary information for these calculations. We will store these files on either NOAA or UAlbany computing systems during and after the project and will make these available to any interested party. The software that will be used to compile and plot the innovation statistics will be available to any UFS or HAFS developer to apply to a different set of forecasts. This project will not generate any other environmental data.

6. REFERENCES


