FORECASTERS’ FORUM

How Do Forecasts Utilize Output from a Convection-Permitting Ensemble Forecast System? Case Study of a High-Impact Precipitation Event

CLARK EVANS
University of Wisconsin–Milwaukee, Milwaukee, Wisconsin

DONALD F. VAN DYKE AND TODD LERICOS*
National Weather Service, Tallahassee, Florida

(Manuscript received 23 May 2013, in final form 18 September 2013)

ABSTRACT
The proliferation of ensemble forecast system output in recent years motivates this investigation into how operational forecasters utilize convection-permitting ensemble forecast system guidance in the forecast preparation process. A 16-member, convection-permitting ensemble forecast of the high-impact heavy precipitation resulting from Tropical Storm Fay (2008) is conducted and evaluated. The ensemble provides a skillful, albeit underdispersive and bimodal, forecast at all precipitation thresholds considered. A forecasting exercise is conducted to evaluate how forecasters utilize the ensemble forecast system guidance. Forecasters made two storm-total accumulated precipitation forecasts: one before and one after evaluating the ensemble guidance. Concurrently, forecasters were presented with questionnaires designed to gauge their thought processes in preparing each of their forecasts. Exercise participants felt that the high-resolution ensemble guidance added value and confidence to their forecasts, although it did not meaningfully reduce forecast uncertainty. Incorporation of the ensemble guidance into the forecast preparation process resulted in a modest mean improvement in forecast skill, with each forecast found to be skillful at all accumulated precipitation thresholds. Forecasters primarily utilized the ensemble guidance to identify a “most likely” forecast outcome from disparate deterministic guidance solutions and to help quantify the uncertainty associated with the forecast. Forecasters preferred ensemble guidance that enabled them to quickly understand the range of solutions provided by the ensemble, particularly over the entirety of the domain. Forecasters were generally aware of the diversity of solutions provided by the ensemble guidance; however, only a select few actively interrogated this information when revising their forecasts and each did so in different ways.

1. Introduction
Meteorological ensembles owe substantial motivation to the findings of Lorenz (1963), who demonstrated that a small change in the initial representation of the atmospheric state can result in comparatively large forecast differences between any two otherwise identical forecasts. Though the precise methods of doing so vary from one ensemble prediction system to another, modern ensembles attempt to quantify part or all of the range of plausible forecast outcomes that may be realized from small changes, comparable to observational uncertainty, in the initial representation of the atmospheric state (e.g., Buizza et al. 2005) and/or other aspects of imperfect numerical weather forecast model systems (e.g., Stensrud et al. 2000; Eckel and Mass 2005). Ensemble guidance can illuminate the uncertainty inherent in a given forecast and in so doing providing probabilistic guidance for desired forecast elements (e.g., Eckel and Mass 2005; Clark et al. 2008). The development, refinement, and increasing utilization of ensemble prediction systems within the atmospheric sciences has led to demonstrated scientific (e.g., Leith 1974; Toth and Kalnay 1997), societal, and economic (Warner 2011 and references therein) benefits.
Rapid advances in computer technology have resulted in an exponential increase in the amount of data available to forecasters and researchers from both deterministic and ensemble modeling systems. However, as the amount of time available to make a forecast in an operational setting has not increased, nor has the amount of data that an individual forecaster or researcher can mentally process in a meaningful fashion when interrogating any given weather event, how ensemble guidance is best utilized remains uncertain. Ongoing experiments on the global scale [e.g., The Observing System Research and Predictability Experiment (THORPEX) Interactive Grand Global Ensemble, TIGGE; Bougeault et al. (2010)] and on the regional scale [e.g., the National Oceanic and Atmospheric Administration (NOAA) Hazardous Weather Testbed Experimental Forecast Program Spring Forecast Experiment (Clark et al. 2012; Kain et al. 2013)] serve, in part, as testbeds for the development of probabilistic multiscale weather forecasting methods.

Forecaster utilization of ensemble guidance is a function of their training and experience, their intuition, and their interpretation of its utility to a given forecast situation. As one might expect, significant differences in one or more of these attributes exist between any two forecasters. Consequently, precisely how forecasters utilize ensemble guidance to prepare and revise forecasts may vary substantially between any two forecasters. Differences in how forecasters utilize deterministic numerical guidance and, in light of its output, prepare a forecast are well known and accepted. It remains an open question, however, as to what these differences are in the context of ensemble-based guidance, why such differences exist, and what the impact of such differences is upon forecast evolution and skill.

To our knowledge, there exist few publications in the peer-reviewed literature that address these or related questions. Novak et al. (2008) reported the results of a survey conducted to comprehensively assess operational forecaster uncertainty needs. They found that many forecasters use ensemble guidance to assess uncertainty and that forecasters desire access to the output from individual ensemble members when assessing forecast uncertainty. However, ensemble underdispersion and bias promote skepticism among some forecasters regarding the value ensemble guidance adds over that provided by higher-resolution deterministic and/or multimodel ensemble guidance. Novak et al. (2008) also found that forecasters desire higher-resolution ensemble prediction systems that are capable of providing uncertainty information for resolution-dependent phenomena such as precipitation bands and mesoscale convective systems. Likewise, the lack of postprocessed ensemble guidance directly tied to end-user products produced by forecasters was shown to be a hindrance to more widespread forecaster acceptance of ensemble guidance.

Forecasters’ impressions of the utility of ensemble guidance to the forecast process are also documented by Rotach et al. (2009) and Clark et al. (2012). Within the context of the Mesoscale Alpine Programme Demonstration of Probabilistic Hydrological and Atmospheric Simulation of Flood Events (MAP D-PHASE), Rotach et al. (2009) found that limited-area ensemble prediction systems had a significant positive impact upon preparing quantitative precipitation forecasts and conveying forecast confidence. However, they also noted that appropriate visualization methods were necessary in order for forecasters to effectively utilize ensemble forecast system output and that forecaster acceptance of ensemble guidance products varied substantially between individual forecasters. Within the context of the NOAA Hazardous Weather Testbed 2010 Spring Forecast Experiment, Clark et al. (2012) reported upon the perceived skill of convection-permitting versus convection-parameterizing ensemble forecast systems for warm-season precipitation forecasts. Subjectively, a convection-permitting ensemble guidance system was found to provide added value over convection-parameterizing ensemble guidance. Such improvement was described as “transformational” in nature, providing motivation toward the development of operational convection-permitting deterministic and ensemble forecast systems.

The primary goal of this forward-looking research is to evaluate how operational forecasters utilize convection-permitting ensemble forecast system guidance when a high-impact weather event is forecast. In so doing, we extend upon the findings of Novak et al. (2008) to focus upon a specific forecast challenge and examine precisely how ensemble guidance—notably, that from a convection-parameterizing ensemble— is used in the forecast process. Relevant questions addressed as a part of this research include, but are not necessarily limited to, the following:

- How do forecasters make use of information related to forecast diversity and spread when making or refining a deterministic forecast?
- How do forecasters evaluate ensemble forecast system dispersion, if they do at all, and how does it impact their use of the ensemble forecast data in the forecast process?
- Do forecasters inherently trust a new piece of forecast guidance until given a reason to not trust it, or do forecasters inherently distrust a new piece of forecast guidance until given a reason to trust it?
- As no two forecasters forecast in an identical manner, what is the nature of the variability in how forecasters utilize ensemble forecast system output?
To help address these and other related questions, a 16-member ensemble forecast of a high-impact heavy precipitation event is conducted and evaluated. A forecasting exercise is developed to enable both objective and subjective assessments of how the ensemble forecast data influence human forecasts of accumulated precipitation for the selected event. The chosen event, ensemble forecast system formulation, and forecasting exercise construction are discussed in section 2. An objective assessment of how incorporating ensemble forecast system output into the forecast process impacts human forecast skill is presented in section 3. Interpretation of feedback obtained from the forecasters during the forecasting exercise is presented in section 4. Concluding remarks are presented in section 5.

2. Methodology

a. Case selection and overview

The heavy precipitation event examined in this study is that associated with Tropical Storm Fay (2008; Brown et al. 2010). Fay was responsible for extreme rainfall, defined here as accumulated precipitation in excess of 250 mm (~10 in.), across Hispaniola, Cuba, Florida, and southern portions of Georgia and Alabama between 16 and 24 August 2008. In the United States alone, the estimated total damage from rainfall-induced flooding was $560 million. In the Tallahassee, Florida, National Weather Service Weather Forecast Office (NWS WFO) County Warning Area (CWA), widespread heavy to extreme rainfall was observed between 22 and 24 August 2008 (Fig. 1). The maximum observed rainfall was at Thomasville, Georgia, where 698.5 mm (27.50 in.) of rainfall fell during this 3-day period (Brown et al. 2010). At Tallahassee, the accumulated precipitation total of 290.5 mm (11.44 in.) during this period ranks as the second-highest 3-day rainfall total between 1948 and 2010. Not surprisingly, widespread major to record flooding resulted from Fay’s storm-total rainfall across the region (e.g., NWS 2012). On the larger scale, the track and intensity of Fay were operationally well forecast (Stewart and Beven 2009). Available large-scale numerical and operational (Fig. 2) and ensemble-based (e.g., Schumacher and Davis 2010) forecast guidance highlighted the potential for locally heavy, albeit perhaps not extreme, rainfall in association with Fay across the southeastern United States. Likewise, the rainfall event was generally considered to be well forecast operationally (D. Novak 2012, personal communication).

b. Ensemble forecast system formulation

Version 3.3 of the Advanced Research core of the Weather Research and Forecasting (WRF-ARW; Skamarock et al. 2008) mesoscale model is used to conduct the numerical simulations that compose the ensemble forecast system. The simulation domain utilized for these experiments contains 500 × 500 × 40 grid points and is centered at 30.4°N, 84.1°W. The horizontal grid spacing for these simulations is 4 km. All simulations begin at 0000 UTC 22 August 2008, 24–36 h prior to the onset of the heaviest precipitation, and end at 0000 UTC 25 August 2008, more than 12 h after the culmination of the heaviest precipitation.

Initial and lateral boundary conditions for the control member of the ensemble forecast system are provided at a horizontal grid spacing of 0.5° from the 0000 UTC 22 August 2008 forecast cycle of the Global Forecast System (GFS) forecast cycle. Physical parameterization formulations utilized by the control simulation include the WRF single-moment six-class (Hong and Lim 2006) microphysical scheme, the Yonsei University (Hong et al. 2006) planetary boundary layer scheme, the Rapid Radiative Transfer Model (RRTM) longwave (Mlawer et al. 1997) radiation model, Dudhia’s (1989) shortwave radiation, the unified Noah land surface model (Chen and Dudhia 2001), the three-category urban canopy model (Chen et al. 2006), and Garratt’s (1992) surface heat and momentum transfer coefficient parameterizations. No convective parameterization is used within the model simulations. This configuration is similar to but slightly updated from the configuration of the Advanced Hurricane WRF model described by Davis et al. (2008).

The ensemble forecast system utilized in this study contains 16 members. Ensemble diversity is achieved
through variability in the initial and lateral boundary conditions, the representation of the sea surface state, the utilization of a stochastic kinetic energy backscatter perturbation parameterization, and the selections of the planetary boundary layer and microphysical parameterizations. Details behind the formulation of each ensemble member are provided in Table 1. This ensemble configuration promotes little diversity between individual members in terms of forecast locations of the heaviest rainfall and modest diversity in rainfall amounts at these locations (e.g., Fig. 3). The former is primarily due to large-scale initial and lateral boundary condition variability whereas the latter is due primarily to fundamental differences between individual physical parameterization packages utilized within the ensemble.

c. Forecasting exercise

To assess how forecasters utilize output from the ensemble forecast system, a forecasting exercise was developed. The forecasting exercise comprised two stages. In the first stage of the exercise, forecasters were presented with the 72-h accumulated precipitation forecast issued at 0000 UTC 22 August 2008 by the Hydrometeorological Prediction Center (HPC) and kinematic and thermodynamic forecast fields from the 0000 UTC 22 August 2008 cycles of the Global Forecast System (GFS), North American Mesoscale (NAM), and European Centre for Medium-Range Weather Forecasts (ECMWF) deterministic models. Utilizing these data and their subjective interpretation of the forecast event, they were asked to make a 72-h deterministic quantitative precipitation forecast (QPF) for the forecast region of interest. By means of comparison, NWS WFO forecasters typically utilize forecasts provided by deterministic and/or ensemble guidance, the HPC, or a subjective guidance blend as a starting point when preparing a QPF. In this exercise, forecasters were free to utilize this approach or to start from scratch as they best saw fit. In the second stage of the exercise, forecasters were presented with output from the ensemble forecast system and asked to revise their initial forecasts. Particular focus was given to providing specialized postprocessed ensemble guidance.

![Fig. 2. Forecast accumulated precipitation (shaded, mm) between 0000 UTC 22 Aug and 0000 UTC 25 Aug 2008 from the 0000 UTC 22 Aug 2008 forecasts from the (a) GFS, (b) NAM, (c) HPC, and (d) ECMWF models.](image-url)
directly tied to the forecast problem—that of accumulated precipitation—at hand.

After completing each of their forecasts, forecasters were presented with questionnaires designed to gauge their thought processes in preparing and revising their forecasts. Prior to being presented with output from the ensemble forecast system, forecasters were asked to

- evaluate the utility of the available guidance in making their forecast;
- describe which deterministic guidance products they felt had the best and worst handle on the forecast situation and why they thought this was the case; and
- describe how they utilize regularly available, predominantly convection-parameterizing ensemble guidance, such as the National Centers for Environmental Prediction’s (NCEP) Global Ensemble Forecast System (GEFS) and Short-Range Ensemble Forecast (SREF), in routine forecast operations.

After revising their forecasts with ensemble forecast system output, forecasters were asked to

- state whether the provided ensemble data added value to their forecast; in other words, did the provided ensemble data enable them to convey, whether implicitly or explicitly, information (e.g., some measure of forecast confidence) that their original forecast did not or was not able to convey?
- provide a subjective assessment of forecast uncertainty given the provided ensemble data;
- describe how strongly the provided ensemble data influenced changes to their forecasts;
- describe how and why the provided ensemble data influenced changes to their forecasts; and
- describe which means of synthesizing ensemble output were most and least useful to them in the forecast process.

These responses are used to subjectively evaluate how forecasters utilize output from the ensemble forecast system.

In total, nine forecasters completed the exercise during March and April 2012. Of the nine forecasters who completed the exercise, seven were from the Tallahassee WFO, one was from the Florida Division of Emergency Management, and one was from the Peachtree City, Georgia, WFO. All forecast exercise results presented herein are presented anonymously. Seven forecasters completed the exercise in its entirety; the other two forecasters completed only portions of the exercise. Coupled with the focus of the exercise on a single high-impact meteorological event, the relatively small number of participants limits the ability to generalize the results presented herein. Further research is planned to assist in the

<table>
<thead>
<tr>
<th>Experiment</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>GFS-Control</td>
<td>Control simulation using 0000 UTC 22 Aug 2008 0.5° GFS forecast data for initial and lateral boundary conditions</td>
</tr>
<tr>
<td>GFS-PBL</td>
<td>As in GFS-Control, except with the Mellor–Yamada–Janjić (Janjić 1994) planetary boundary layer parameterization</td>
</tr>
<tr>
<td>GFS-MP</td>
<td>As in GFS-Control, but with the Thompson et al. (2008) microphysical parameterization</td>
</tr>
<tr>
<td>GFS-SST</td>
<td>As in GFS-Control, but using the 0000 UTC 22 Aug 2008 0.083° RTG_SST_HR (<a href="http://polar.ncep.noaa.gov/sst/ophi/">http://polar.ncep.noaa.gov/sst/ophi/</a>) sea surface temperature analysis</td>
</tr>
<tr>
<td>GFS-Stochastic</td>
<td>As in GFS-Control, but also employing the stochastic kinetic energy backscatter (Berner et al. 2011) perturbation scheme</td>
</tr>
<tr>
<td>GFS-PBL+SST</td>
<td>As in GFS-Control, but utilizing both the GFS-PBL and GFS-SST modifications</td>
</tr>
<tr>
<td>GFS-PBL+MP+SStochatic</td>
<td>As in GFS-Control, but utilizing the GFS-PBL, GFS-MP, and GFS-Stochastic modifications</td>
</tr>
<tr>
<td>NAM-Control</td>
<td>Control simulation using 0000 UTC 22 Aug 2008 0.2° NAM model forecast data for initial and lateral boundary conditions</td>
</tr>
<tr>
<td>NAM-PBL</td>
<td>As in NAM-Control, but with the Mellor–Yamada–Janjić (Janjić 1994) planetary boundary layer parameterization</td>
</tr>
<tr>
<td>NAM-MP</td>
<td>As in NAM-Control, but with the Thompson et al. (2008) microphysical parameterization</td>
</tr>
<tr>
<td>NAM-SST</td>
<td>As in NAM-Control, but using the 0000 UTC 22 Aug 2008 0.083° RTG_SST_HR (<a href="http://polar.ncep.noaa.gov/sst/ophi/">http://polar.ncep.noaa.gov/sst/ophi/</a>) sea surface temperature analysis</td>
</tr>
<tr>
<td>NAM-Stochastic</td>
<td>As in NAM-Control, but also employing the stochastic kinetic energy backscatter (Berner et al. 2011) perturbation scheme</td>
</tr>
<tr>
<td>NAM-PBL+SST</td>
<td>As in NAM-Control, but utilizing both the NAM-PBL and NAM-SST modifications</td>
</tr>
<tr>
<td>NAM-PBL+MP+SStochatic</td>
<td>As in NAM-Control, but utilizing the NAM-PBL, NAM-MP, and NAM-Stochastic modifications</td>
</tr>
<tr>
<td>NAM-PBL+SST+SStochatic</td>
<td>As in NAM-Control, but utilizing the NAM-PBL, NAM-SST, and NAM-Stochastic modifications</td>
</tr>
</tbody>
</table>
FIG. 3. The 72-h forecast accumulated precipitation (shaded, mm), valid at 0000 UTC 25 Aug 2008, from the (a)–(p) 16 members of the ensemble forecast system in the order listed in Table 1, (q) ensemble mean, (r) maximum of all ensemble members, and (s) minimum of all ensemble members.
generalization of these results to larger sets of meteorological events, whether well forecast (such as Fay) or poorly forecast, and/or human forecasters.

Note that there exists significant latent “forecaster memory” of Tropical Storm Fay (2008) among forecasters in the local WFO given its substantial impacts to the region. To mitigate potentially deleterious impacts of such memory upon the exercise, the forecast exercise was construed as Tropical Storm Trixie, a fictitious tropical storm impacting the Houston–Galveston, Texas, metropolitan area. To achieve this, prior to their being presented to forecasters, gridded forecast products for Tropical Storm Fay (2008) from the 0000 UTC 22 August 2008 forecast cycle were shifted south and west by 1° latitude and 11° longitude, respectively, to shift the forecast region of interest from the Tallahassee WFO CWA to the Houston–Galveston WFO CWA. An illustrative example of this shift is presented in Fig. 4. No specific date information was provided to forecasters during the exercise. For reference, a complete listing of forecast products provided for both stages of the forecasting exercise may be found in the appendix.

3. Results: Objective analysis

a. Assessment of ensemble forecast system skill and dispersion

To assess the underlying skill of the ensemble forecast system, 72-h accumulated precipitation forecasts valid at 0000 UTC 25 August 2008 from the ensemble forecast system are verified against 6-hourly NCEP stage IV multisensor accumulated precipitation observations (Lin and Mitchell 2005). The spatial domain over which the verification is conducted approximates the Tallahassee WFO CWA and encompasses a 4° latitude × 5° longitude domain centered on Tallahassee. As the native horizontal grid spacing of the stage IV data (Δx ≈ 4.75 km) is slightly coarser than that of the ensemble forecast system (Δx = 4 km), ensemble forecast system data are linearly extrapolated to the horizontal grid of the stage IV data prior to forecast verification.

The forecast skill of the ensemble is assessed by computing the area under the receiver operating characteristic (ROC) curve (e.g., Hamill and Juras 2006). A ROC curve relates the hit rate to the false alarm rate for a series of ranked ensemble forecasts. A perfect forecast has an area under the ROC curve equal to 1, whereas a forecast that is said to be randomly drawn from climatology has an area under the ROC curve equal to 0.5. The area under the ROC curve is computed at each of 10 accumulated precipitation thresholds: 10, 25, 50, 75, 100, 150, 200, 250, 300, and 350 mm. Figure 5 depicts the area under the ROC curve for the full 16-member ensemble forecast system. The ensemble provides a skillful forecast at all precipitation thresholds with particularly enhanced skill (as compared to climatology) noted at the lower thresholds. For comparison, these values are somewhat lower than those obtained from a related analysis by

![Fig. 4. (a) Ensemble mean accumulated precipitation (shaded, mm) between 0000 UTC 22 Aug and 0000 UTC 25 Aug 2008. (b) As in (a), but for the domain shifted westward by 11° longitude and southward by 1° latitude.](image-url)
Schumacher and Davis (2010) that utilized ECMWF Ensemble Prediction System forecast data. It should be noted, however, that their verification utilized coarser \((D = 0.5^\circ)\) forecast and observational data and considered accumulated precipitation over both a longer time period (120 h) and a larger domain (central and eastern United States) than in this work.

In Fig. 5, the area under the ROC curve is also depicted for two 8-member subsets of the full 16-member ensemble: the GFS- and NAM-based ensemble members. The GFS-based ensemble members are skillful at all accumulated precipitation thresholds. Conversely, the NAM-based ensemble members are only slightly more skillful than climatology at the 150-mm accumulated precipitation threshold and lower and actually degrade the performance of the full 16-member ensemble at all thresholds. This largely reflects characteristics, albeit downscaled and perturbed, of the large-scale deterministic forecasts that provided initial and lateral boundary conditions for these simulations (cf. Figs. 1 and 2a,b). Consequently, it is unsurprising that ensemble forecast system skill is comparable to that of the available deterministic forecast guidance at the 200-mm accumulated precipitation threshold and lower (not shown). It is at higher thresholds, which were forecast by none of the available deterministic guidance (e.g., Fig. 2), that the ensemble forecast system—and, in particular, its GFS-initialized members—add limited value over climatology to the forecast by explicitly highlighting the potential for a higher-end rainfall event.

The dispersive nature of the ensemble is evaluated utilizing a rank histogram approach (e.g., Wilks 2011). A rank histogram compares ranked ensemble forecasts to their verifying observations on a common grid. At each grid point, the verifying observation falls into one of \(n + 1\) bins, where \(n\) is the total number of ensemble members. Bin 1 represents the percentage of verifying observations with a value less than the lowest-ranked ensemble member for that field; conversely, bin \(n + 1\) represents the percentage of verifying observations with a value greater than the highest-ranked ensemble member for that field.

Figure 6 depicts the rank histogram for 72-h accumulated precipitation forecasts verifying at 0000 UTC 25 August 2008 from the full 16-member ensemble. Black bars represent the rank histogram for the entirety of the ensemble simulation domain, whereas gray bars represent the rank histogram only for the 4° latitude \(\times 5^\circ\) longitude region described above. For each of the two distributions depicted in Fig. 6, peaks are found in bins 1, 9 \([n/2] + 1\), and 17 \(n + 1\). This implies that the ensemble system is overconfident that the verifying observation will fall within either the range of forecasts provided by the NAM-based ensemble members (which generally represent ranked members 1–8) or the range of forecasts provided by the GFS-based ensemble.

![Fig. 5. Area under the ROC curve for 72-h accumulated precipitation thresholds of 10, 25, 50, 75, 100, 150, 200, 250, 300, and 350 mm for the full 16-member ensemble (black line), the 8 GFS-based ensemble members only (dark gray line), and the 8 NAM-based ensemble members only (light gray line).](image-url)
members (which generally represent ranked members 9–16) with few forecasts (but many observations) outside of each range of forecasts. The underdispersive nature of such an ensemble was also highlighted by Clark et al. (2008, 2009) in their examinations of precipitation forecasts from both convection-parameterizing and convection-permitting ensembles. However, it should be noted that the ensemble system considered herein is notably less dispersive than those considered by Clark et al. (2008, 2009). The underdispersive nature of ensemble forecast systems, whether convection parameterizing or convection permitting, has been addressed by Eckel and Mass (2005) and Novak et al. (2008), among others.

Given the presence of the synoptic- to meso- scale attractor provided by Tropical Storm Fay, the severe underdispersion of our ensemble forecast system is likely due in large part to the minimal initial condition and lateral boundary condition diversity that it employs, particularly at later forecast times where synoptic-scale error growth becomes nonlinear (e.g., Gilmour et al. 2001). While the forecast attractor provided by the GFS-based ensemble members reasonably resembles the “true” forecast attractor, the ensemble provides insufficient bounds upon this attractor. Variance in model physics, along with the utilization of the stochastic kinetic energy backscatter scheme of Berner et al. (2011) and varying the initial sea surface temperature dataset, contributes to ensemble dispersion, but of smaller magnitude and primarily on smaller scales. These issues are not uniquely identified with our research, however (e.g., Eckel and Mass 2005; Clark et al. 2008, 2009, among others), although we again note that our ensemble is less dispersive than their underdispersive ensembles. However, we argue that the relative lack of ensemble dispersion does not compromise our ability to draw sound insight into the key questions motivating our research, namely to better understand how forecasters make use of information related to forecast diversity and spread when making or refining a forecast, whether they assess ensemble dispersion and how it impacts their use of the ensemble data, and whether they inherently trust or distrust a new piece of forecast guidance.

In the aggregate, the skill exhibited by the convection-permitting ensemble forecast system, particularly at the highest precipitation thresholds, suggests that it is capable of producing skillful, if underdispersive and somewhat biased, forecasts of heavy to extreme rainfall for the specific case and forecast cycle considered herein. Further research is necessary, however, to examine whether

**Fig. 6.** Rank histogram evaluating the dispersion of 72-h accumulated precipitation (mm) ensemble member forecasts as compared to 72-h NCEP stage IV accumulated precipitation (mm) observations. Black bars represent the rank histogram over the entirety of the ensemble simulation domain, whereas gray bars represent the rank histogram only over the 4° latitude × 5° longitude verification region described in section 3.
these statements can be generalized to other forecast cycles and/or heavy precipitation events.

b. Overview of exercise participant forecasts

To a large extent, individual participant forecasts made after stage 1 of the forecasting exercise resemble one or more of the deterministic guidance products (cf. Figs. 2 and 7a–g), though there exists substantial variability between individual participant forecasts. Notably, the maximum forecast accumulated precipitation is generally higher in participant forecasts than in the deterministic guidance. The consensus forecast (Fig. 7h), obtained as the mean of the seven participants’ forecasts, has a maximum of approximately 350 mm in the eastern Florida panhandle, in close proximity to where the maximum precipitation fell (Fig. 1).

Individual participant forecasts made after stage 2 of the forecasting exercise bear many similarities to those made after stage 1 of the forecasting exercise (cf. Figs. 7 and 8). However, upon further inspection, subtle differences emerge in how each participant made use of the ensemble forecast guidance to update their initial forecasts. First, in general, the ensemble guidance did not lead participants to increase their overall maximum forecast rainfall totals. Instead, forecast changes primarily focused upon the location and areal extent of the heaviest accumulated precipitation totals.

One forecaster, forecaster F, declined to modify their forecast in light of the ensemble guidance (Fig. 9f), the reasoning for which is discussed in section 4. Two forecasters, forecasters A and C, made relatively minor changes to their initial forecasts, reducing forecast precipitation totals in the western portion of the domain and increasing forecast precipitation totals in the eastern portion of the domain (Figs. 9a,c). Four forecasters, forecasters B, D, E, and G, made more substantial changes to their initial forecasts (Figs. 9b,d,e,g). Forecaster B eliminated their southwestern rainfall maximum,
forecaster D expanded their rainfall maximum to the south and east, forecaster C cut back on their forecast precipitation in the southern half of the forecast domain, and forecaster G shifted their rainfall maximum eastward and expanded its areal extent. Changes to each individual forecast manifest themselves in small changes to the consensus forecast with generally increased (decreased) forecast accumulated rainfall in the northeastern (southwestern) portion of the forecast domain (Fig. 9h).

c. Impact of ensemble forecast system output on human forecaster skill

The impact upon forecast skill due to the incorporation of ensemble forecast system output into the forecast is assessed by verifying both initial and revised forecasts against accumulated precipitation estimates from 6-hourly NCEP stage IV precipitation analyses. All verification of exercise participant forecasts is conducted on the stage IV grid with $\Delta x \approx 4.75$ km. For verification purposes, initially hand-drawn forecasts are digitized utilizing the Graphical Forecast Editor (GFE) localized to the Houston–Galveston WFO. The digitized forecasts are subsequently extracted to GFE-standard Network Common Data Form (netCDF) files and shifted northward by 1° latitude and eastward by 11° longitude (the inverse of that for the forecasting exercise) for verification and further analysis.

From this verification, several gridpoint contingency-table-based skill metrics are computed at each of 10 accumulated precipitation thresholds: 10, 25, 50, 75, 100, 150, 200, 250, 300, and 350 mm. The metrics considered here include the equitable threat score (ETS), Heidke skill score (HSS), and Kuipers skill score (KSS). The ETS provides a measure of how well forecasts compare to observations when accounting for random chance. Allowable values of the ETS range from $\frac{1}{3}$ to 1, with 0 indicating no skill and 1 indicating a perfect score. The HSS is similar to the ETS, providing a measure of the percent improvement of the forecast as compared to pure chance. Allowable values of the HSS range from $-\infty$ to 1, with 0 indicating no skill and 1 indicating a perfect score. Finally, the KSS provides a measure of how well separated accurate forecasts are from falsely positive forecasts. Allowable values of the KSS range from $-1$ to 1, with 0 indicating no skill and 1 indicating...
a perfect score. More details regarding the computation and interpretation of these metrics are provided by Wilks (2011) and the Centre for Australian Weather and Climate Research (2012).

Skill scores for participant forecasts completed at the end of stage 2 of the forecasting exercise are presented in Fig. 10. Participant forecasts are skillful at all thresholds for which precipitation was forecast. Despite the differences between their respective forecasts, there exist similar levels of forecast skill between each of the seven participants’ forecasts. For all forecasts, the KSS is at or above 0.4, and sometimes substantially so, at all precipitation thresholds for which precipitation was forecast. The HSS and ETS are highest at the lower precipitation thresholds and gradually decay with increasing forecast accumulated precipitation totals. All exercise participants' forecasts exhibit greater skill than available deterministic and operational forecasts issued at 0000 UTC 22 August 2008 (cf. Figs. 10 and 11).

The mean skill score improvement realized between stages 1 and 2 of the forecasting exercise is depicted in Fig. 12. In the mean, forecasts completed after the interrogation of ensemble forecast system output are more skillful, in terms of the three skill metrics presented herein, than forecasts completed before the interrogation of ensemble forecast system output. The mean improvement in the ETS and HSS is between 0.05 and 0.1 at all accumulated precipitation thresholds. The mean improvement in the KSS is less than 0.1 at all accumulated precipitation thresholds below 250 mm, above which greater mean improvement is noted.

Individual forecaster skill score changes between stages 1 and 2 of the forecasting exercise are depicted in Fig. 13. As might be expected, there is a fair amount of variability in skill score changes between individual forecasters. The greatest changes in skill scores are noted with forecasters B, D, E, and G, or the forecasts that changed the most between stages 1 and 2 of the

FIG. 9. The 72-h post–stage 1 minus post–stage 2 accumulated precipitation forecasts (shaded, mm) for (a)–(g) each of the seven forecasters who completed the forecasting exercise in its entirety and (h) the consensus forecast. As in Fig. 7, but note that all forecasts were initially prepared on the geographically shifted forecast domain over the Houston–Galveston WFO CWA described in section 2c and subsequently shifted eastward and northward for verification and visualization.
forecasting exercise (Fig. 9). However, the skill of these forecasts is comparable to that of those produced by forecasters whose forecasts changed by lesser amounts after interrogating ensemble forecast system output.

We note that the limited sample size precludes statistical significance testing upon the differences presented in Fig. 12. This, coupled with the limited extent of the exercise, hinders the generality of insight that can be drawn from the above-presented analyses but provides substantial motivation for further research to better quantify these findings.

4. Results: Subjective analysis

a. Post–stage 1 findings

After creating an initial accumulated precipitation forecast utilizing only deterministic and national center guidance, forecasters were asked a series of questions, as listed in section 2c. Even with a limited sample size of nine participants, a rich variety of answers to each question was provided by the participants. With this in mind, we focus on illustrating common themes from the survey responses. First, with an average response of 3.33 out of 5, exercise participants felt that the available deterministic and national center guidance was at least somewhat useful to them with regard to forecasting a QPF total (Table 2). No forecasters felt that it was not useful, while one forecaster (forecaster B) thought that it was extremely useful.

Participant feedback describing which deterministic guidance products they felt had the best and worst capability to handle the forecast situation are provided in Table 2. In general, forecasters interpreted the model QPF in light of how it compared to the official National Hurricane Center (NHC) forecast track of the tropical
cyclone. However, as the survey responses illustrate, there is notable variability among exercise participants in how they interpreted the available guidance. This is best illustrated through an examination of the survey responses that indicated the NAM and ECMWF model QPF forecasts as the most and/or least likely to be accurate. Of the three forecasters who believed that the NAM model QPF was least likely to be accurate, forecasters C and F indicated that the NAM’s relatively poor performance (as compared to global deterministic guidance) with tropical cyclones weighed heavily into their decision-making process. Forecasters D and F further noted that the NAM forecast track of the tropical cyclone was to the south of the official NHC track, further casting doubt in their minds as to the veracity of its track forecast and QPF.

Interestingly, despite their assessment of the ECMWF model QPF as the most likely to be accurate, these three forecasters expressed that the ECMWF model QPF was likely “underdone” (forecaster C), in need of “manual adjustment upward” (forecaster D), and “on the low side” (forecaster F). This concern was expressed by four of the five forecasters who believed that the ECMWF model QPF was least likely to be accurate (forecasters A, B, G, and H), even if they felt other aspects of the ECMWF’s model solution were reasonable. Consequently, forecasters appeared to focus upon identifying either (a) the most reasonable QPF amounts or (b) the most reasonable simulated track from the available guidance. In the former case, the forecast locations of these QPF amounts were modified in light of the participants’ interpretation of the guidance and forecast

![Figure 11](https://example.com/fig11.png)
scenario. In the latter case, the QPF amounts themselves were modified in light of the participants’ interpretation of the guidance and forecast scenario. It should be emphasized, however, that one philosophy is not necessarily better than the other and that they can result in similar forecasts, as illustrated by Fig. 7. Nevertheless, these differences illustrate that human bias, independent of how it arises, influences forecast preparation.

Eight of nine forecasters stated that they utilize regularly available ensemble guidance in routine forecast operations. The remaining forecaster, forecaster F, stated that they are “more likely to use the SREF guidance when a potential big event is in the forecast.” Three of the eight forecasters answering in the affirmative (A, D, E) noted that they utilize regularly available ensemble guidance particularly when refining forecasts at later forecast periods (i.e., beyond 3–5 days). In addition, three forecasters (C, D, I) noted that they make particular use of specialized postprocessed products from regularly available ensemble guidance, focusing upon aviation (e.g., fog, visibility) and severe weather forecasting. The most common themes utilized by forecasters in describing how and why they utilize regularly available ensemble guidance in routine forecast operations were “uncertainty,” “variability,” and “consensus.” Nearly uniformly, forecasters utilize ensemble guidance in routine forecast operations in an attempt to identify the “most likely” forecast outcome from often disparate deterministic guidance solutions and to quantify the uncertainty inherent to the forecast problem at hand.

b. Post-stage 2 findings

After revising their accumulated precipitation forecasts in light of the provided ensemble guidance, forecasters were asked a series of questions, as listed in section 2c. The goal of these questions was to subjectively evaluate how forecasters utilized output from the ensemble forecast system in the forecast preparation process. Herein, we focus on illustrating common themes from the survey responses, tying these themes to their forecasts (Figs. 8 and 9) as appropriate. With an average score of 4 out of 5, exercise participants felt that the provided high-resolution ensemble guidance was useful to them with regard to revising their forecasts (Table 3). No forecasters thought that the ensemble guidance was not useful, while four forecasters (B, C, D, G) thought that the ensemble guidance was extremely useful. Of these four forecasters, three (B, D, G) greatly modified their accumulated precipitation forecasts (Fig. 9), with such modifications generally contributing to increased forecast skill (Fig. 13).

The only forecaster to state that the ensemble guidance did not add value to their forecast was forecaster F (Table 3). Forecaster F stated that “it is apparent that the ensemble approaches have to be used cautiously due to the potential for ensemble artifacts to impact the final outcome.” These ensemble artifacts, or perceived shortcomings in the ensemble configuration, became apparent to this forecaster through “the bi-modal accumulated precipitation distribution,” which they hypothesize is “an artifact of the NAM vs. GFS initial conditions.” Similarly, forecaster I stated that “the ensemble did not seem to be too effective and I somewhat disregarded it because of the severity of the differences of the NAM and the GFS.” Among forecasters who felt that the ensemble guidance added value to their forecast, forecasters A (“it became obvious that there were two separate camps amongst the various ensemble members”), B (“my QPF forecast [was] based on a blend of the GFS and NAM and the GFS and NAM ensemble forecasts allowed me to further compare the similarities and differences with higher detail”), and E (“the ensemble forecasts also depicted two distinct areas of potential higher rainfall totals”) made similar remarks.

These findings indicate that a subset of forecasters attempt to visually interpret ensemble dispersion when evaluating ensemble guidance. However, such forecasters make use of this information in distinctly different ways and, by extension, assign different levels of trust to a new piece of forecast guidance. For forecasters F, who does not regularly incorporate ensemble guidance into routine forecast operations, and I, who makes routine use of ensemble guidance primarily for specific
forecast applications, a perception of a biased or under-disperse ensemble caused them to nearly discard it altogether, following Novak et al. (2008). The same cannot be said for forecasters A, B, and E. Forecaster A noted that the ensemble “added value in communicating the uncertainty in inland portions of Texas and argued for a less robust forecast,” while forecaster E felt that the ensemble “helped me set some bounds” upon the

**Fig. 13.** Difference in ETS (dark solid), HSS (light dashed), and KSS (dotted) scores as a function of precipitation threshold (mm) for 72-h accumulated precipitation forecasts completed after stages 1 and 2 of the forecaster exercise by (a)–(g) each of the seven forecasters who completed the forecasting exercise in its entirety and (h) the consensus forecast. Positive values denote greater skill score values after stage 2 of the forecasting exercise.

**Table 2.** Responses to selected survey questions asked of forecast exercise participants after stage 1, or the “preensemble” phase, of the forecasting exercise. The scale of 1–5 in the first question is defined with 1 meaning not at all useful and 5 meaning extremely useful. Forecasters A–G correspond to the forecasts depicted within panels a–g of Figs. 7–10 and 13, while forecasters H and I are the responses received from two forecasters who did not provide accumulated precipitation forecasts after each stage of the forecasting exercise.

<table>
<thead>
<tr>
<th></th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>E</th>
<th>F</th>
<th>G</th>
<th>H</th>
<th>I</th>
</tr>
</thead>
<tbody>
<tr>
<td>On a scale of 1–5, how useful was the available guidance with respect to forecasting a QPF total?</td>
<td>2</td>
<td>5</td>
<td>4</td>
<td>3</td>
<td>3</td>
<td>2</td>
<td>3</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>Which model QPF did you believe was most likely to be accurate?</td>
<td>GFS</td>
<td>NAM</td>
<td>ECMWF</td>
<td>ECMWF</td>
<td>GFS</td>
<td>ECMWF</td>
<td>NAM</td>
<td>NAM</td>
<td>NAM</td>
</tr>
<tr>
<td>Which model QPF did you believe was least likely to be accurate?</td>
<td>ECMWF</td>
<td>ECMWF</td>
<td>NAM</td>
<td>NAM</td>
<td>ECMWF</td>
<td>NAM</td>
<td>ECMWF</td>
<td>ECMWF</td>
<td>GFS</td>
</tr>
</tbody>
</table>
forecast. Forecaster B noted that the ensemble “provided more confidence in my original solution, especially in the areas of max QPF” and that, “no matter which track verifies best,” the rainfall-related impacts will be the same across the region.

In addition to forecaster B, a number of forecasters indicated that the ensemble guidance added confidence to their forecast. Forecaster C felt that “the ECMWF was the best of the models, and the ensemble charts appeared to confirm this.” Consequently, “I felt more confident in forecasting extreme rainfall totals and in showing more detail in my forecast.” Forecaster D stated that “the mean/min/max QPF fields influenced me a great deal as these values showed a resultant pattern that was similar to my previous forecast that used the ECMWF forecast as a compromise. This increased confidence in forecasting higher QPF totals.” Likewise, forecaster G “was more confident with my higher amounts along with extending more QPF a bit further north than previously.” Despite this and the perceived utility of the ensemble guidance, however, exercise participants felt that the forecast was still modestly uncertain (Table 3). In light of the forecast event, this is not altogether unsurprising. As noted in section 2a, the extreme rainfall realized from Tropical Storm Fay was high impact and locally rare, implying an intrinsically lower degree of predictability. Likewise, the diversity among the guidance in terms of both maximum QPF amount and placement (e.g., Figs. 2 and 3) suggests that uncertainty should remain even if forecast confidence improved.

With an average score of 2.78 out of 5, exercise participants felt that the ensemble guidance only slightly influenced changes to their forecasts (Table 3). This stands in contrast to the perceived utility of the ensemble guidance, described above, and the forecast changes realized between stages 1 and 2 of the forecast exercise (Fig. 9). Forecaster F, who declined to revise their forecast in light of the ensemble guidance, provided the lowest-possible score of 1 in response to this question. Forecasters B and C provided the second-lowest responses, while forecasters E and H provided the highest responses. As noted above, a number of forecasters felt that the ensemble guidance increased confidence in their forecast, independent of the actually realized changes to their forecasts. This indicates that the increased forecast confidence led them to subjectively evaluate changes to their forecasts as being relatively minor in nature. By contrast, forecaster E stated that “it showed me that the operational GFS numbers I went closer to originally were closer to the max ensemble forecast than the mean,” causing them to lower the extent of their higher QPF totals. Forecaster H stated that “it changed my mind as to which model was preferable,” causing them to favor the GFS over the NAM QPF output.

Forecast precipitation information from the ensemble was presented to forecast exercise participants primarily

<table>
<thead>
<tr>
<th>On a scale of 1–5, how useful was the ensemble QPF-focused guidance?</th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>E</th>
<th>F</th>
<th>G</th>
<th>H</th>
<th>I</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>4</td>
<td>2</td>
<td>5</td>
<td>4</td>
<td>2</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Did the ensemble guidance add value to your forecast?</th>
<th>Yes</th>
<th>Yes</th>
<th>Yes</th>
<th>Not sure</th>
<th>Yes</th>
<th>No</th>
<th>Yes</th>
<th>Yes</th>
<th>Not sure</th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>On a scale of 1–5, how uncertain is your ensemble-modified QPF forecast?</th>
<th>4</th>
<th>2</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>3</th>
<th>3</th>
<th>4</th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>On a scale of 1–5, how much did the ensemble guidance influence your forecast?</th>
<th>Matrix</th>
<th>Plume</th>
<th>Mean</th>
<th>Max–min</th>
<th>Mean</th>
<th>Mean</th>
<th>Max–min</th>
<th>Mean</th>
<th>Mean</th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>What method of displaying ensemble data was most useful?</th>
<th>Meteograms</th>
<th>Meteograms</th>
<th>Meteograms</th>
<th>Matrix</th>
<th>Matrix</th>
<th>Meteograms</th>
<th>Plume</th>
<th>Meteograms</th>
<th>Matrix</th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>What method of displaying ensemble data was least useful?</th>
<th>Matrix</th>
<th>Meteograms</th>
<th>Meteograms</th>
<th>Matrix</th>
<th>Matrix</th>
<th>Meteograms</th>
<th>Plume</th>
<th>Meteograms</th>
<th>Matrix</th>
</tr>
</thead>
</table>
in five forms, as described in the appendix. At the conclusion of the forecasting exercise, forecasters were asked to state which methods of displaying ensemble data were most useful and least useful to them in developing their 0–72-h accumulated precipitation forecast, the results of which are depicted in Table 3. The meteograms, matrix charts, and plume charts each provided information related to the ensemble mean accumulated precipitation forecast or the spread about that ensemble mean at specific locations. Conversely, the maximum, minimum, and mean accumulated precipitation charts provided spatial information that established bounds upon the forecast. This implies that forecasters preferred ensemble output that helps them to quickly understand the range of solutions that it provides over specific, detailed information illuminating ensemble spread that can become overwhelming or may not have direct applicability to the current forecast. This finding is supported by insight from the Air Force Weather Agency, who noted that matching their ensemble forecast system’s outputs to specific forecast requirements positively influenced forecaster acceptance and utilization of the ensemble data (E. Kuchera 2012, personal communication).

5. Conclusions

The primary goal of this research was to evaluate how operational forecasters, such as those found within an NWS WFO, utilize high-resolution ensemble forecast system guidance when a high-impact weather event is forecast. To that end, a 16-member, 72-h ensemble forecast of a high-impact heavy precipitation event, Tropical Storm Fay (2008) across southwest Georgia and the Florida panhandle, was conducted. As assessed via the area under the ROC curve, the ensemble provided a highly skillful forecast at all accumulated precipitation thresholds up to 350 mm. However, the ensemble was also found to be underdispersive, exhibiting overconfidence that the verifying precipitation observations would fall within the range of solutions provided by either the eight NAM-based or the eight GFS-based ensemble members.

To evaluate how forecasters utilized output from the ensemble forecast system, a forecasting exercise was developed. Nine forecasters participated in the forecasting exercise, which was conveyed to exercise participants as a fictitious Tropical Storm Trixie impacting southeast Texas and southwest Louisiana. In the first stage of the exercise, forecasters were presented with a standard suite of operational and gridded numerical forecast products. Utilizing these data and their subjective interpretation of the forecast event, they were asked to make a 72-h accumulated precipitation forecast for the geographically shifted forecast region of interest. In the second stage of the exercise, forecasters were presented with output from the ensemble forecast system and asked to revise their initial forecasts. After completing each stage of the exercise, forecasters were presented with questionnaires designed to gauge their thought processes in preparing and revising their forecasts.

Objectively, individual participant forecasts made after the first stage of the forecasting exercise largely resembled one or more of the deterministic guidance products, albeit with substantial variability noted between individual participant forecasts. Ensemble guidance influences upon these forecasts manifest themselves primarily in terms of the location and areal extent of the heaviest accumulated precipitation totals, with subtle differences noted between individual forecasters in how the ensemble guidance was utilized to revise their initial accumulated precipitation forecasts. Revised accumulated precipitation forecasts were found to be skillful at all thresholds for which precipitation was forecast, with fairly uniform skill noted between individual participant forecasts despite the magnitude of the influence of the ensemble guidance upon their forecasts. Incorporation of the ensemble forecast guidance resulted in a modest mean skill score improvement for each of the three skill metrics considered in this work, albeit with a fair amount of variability noted between individual participants’ forecasts.

In routine forecast operations, exercise participants utilize ensemble guidance to aid in identifying a “most likely” forecast outcome from often disparate deterministic guidance solutions and to help quantify the uncertainty inherent in the given forecast situation. Particular emphases are given to utilizing postprocessed products from regularly available ensemble guidance, particularly products that have utility for high-impact meteorological phenomena, and to evaluating ensemble diversity and spread for forecasts valid at lead times of 3 days or greater. For the specific case examined herein, with few exceptions, forecasters preferred ensemble forecast guidance that enabled them to quickly understand the range of solutions provided by the ensemble over specific, detailed information illuminating ensemble spread that could become overwhelming or may not have had direct applicability to the specific forecast challenge at hand. This both supports (e.g., direct applicability of guidance to the forecast) and differs from (e.g., amount of detail used) the findings of Novak et al. (2008); however, it should be noted that this latter point does support the findings of Rotach et al. (2009).

A subset of forecasters attempted to visually interpret ensemble dispersion when evaluating the provided ensemble guidance. However, each forecaster made use of
this information in distinctly different ways. The perceived underdispersive or biased character of the ensemble led two exercise participants to nearly discard the ensemble guidance altogether, similar to the concerns expressed in Novak et al. (2008). Separately, these two forecasters stated that they make minimal use of ensemble guidance in routine forecast operations. Conversely, three forecasters identified the ensemble forecast solutions as bimodal in nature, yet each felt that it added value to their forecast in terms of communicating uncertainty, setting upper and lower forecast bounds, and adding confidence to their forecasts. These differences, in part, imply that substantial variability exists in how forecasters evaluate the utility of a given piece of guidance, such as from an ensemble and, consequently, weight it when preparing a forecast. It remains an open, yet difficult to answer, question as to how well forecasters make use of ensemble guidance in the forecast preparation process.

Differences in forecast philosophy emerged prior to the interrogation of ensemble forecast system output. Forecasters interpreted deterministic accumulated precipitation guidance in light of how it compared to the official NHC forecast track of the tropical cyclone. Forecasters focused upon identifying either the most reasonable QPF amounts or the most reasonable simulated tropical cyclone track from the range of provided deterministic guidance solutions. In the former case, forecasters subsequently modified the forecast locations of these QPF amounts, whereas in the latter case, forecasters subsequently modified the QPF amounts themselves. As each forecast philosophy resulted in similar accumulated precipitation forecasts, however, neither exhibited greater skill for this particular event. Exercise participants felt that the high-resolution ensemble guidance was useful to them as they revised their initial accumulated precipitation forecasts. A direct connection was noted between those who felt most positively about its utility and the amount that their accumulated precipitation forecasts changed in light of the ensemble guidance. Despite differences in how the ensemble guidance was utilized to revise their accumulated precipitation forecasts, however, most exercise participants felt that it added value and confidence to their forecasts even as a modest degree of forecast uncertainty remained.

While enlightening, the generality of insight that can be drawn from the forecasting exercise detailed herein is somewhat limited given the small sample sizes of forecasters, forecast lead times, ensemble configurations, and meteorological phenomena examined in this work. The limited sample size of forecasters precludes statistical significance testing from being conducted upon both objective and subjective findings presented in this work. Likewise, the applicability of the findings presented herein to present forecast operations is somewhat limited as convection-permitting ensemble guidance, such as the Storm-Scale Ensemble of Opportunity (e.g., Clark et al. 2012), has only recently become available (in limited form) to local WFOs. Consequently, further research is necessary to better understand how operational forecasters utilize ensemble guidance in the operational forecast process. Investigations of this nature have the potential to help target ongoing research-to-operations efforts aimed at developing robust, easily accessible postprocessed ensemble guidance that may be used to target and improve specific elements of both short- and long-range forecasts (e.g., Clark et al. 2012).

Acknowledgments. We are greatly indebted to the nine National Weather Service and Florida Division of Emergency Management forecasters who participated in the forecasting exercise detailed herein. Without their participation, this work would not have been possible. This research benefited from discussions with James Correia, Rich Grumm, Evan Kuchera, and Paul Roebber, as well as reviews by Russ Schumacher and two anonymous reviewers. Numerical simulations carried out in this research were conducted on the avi supercomputing facility at the University of Wisconsin-Milwaukee. Archived GFS, NAM, and HPC forecast products were obtained from the NCDC. Archived ECMWF forecast products were obtained from the UCAR Research Data Archive. Historical precipitation data for Tallahassee, Florida, were obtained from the Utah Climate Center. Forecast exercise questionnaires were hosted on the Google Docs platform. The Model Evaluation Toolkit was used to compute the rank histogram of the ensemble forecast system. This research was partially supported by Subaward Z11-91832 with UCAR under the sponsorship of NOAA as part of the COMET Outreach Program.

APPENDIX

Forecasting Exercise Data

For the first stage of the forecasting exercise, forecasters are provided with a diverse range of gridded meteorological forecast data. From the GFS, NAM, and ECMWF deterministic forecasts, forecasters are provided with forecast charts of 2-m dewpoint temperature (°F); 2-m air temperature (°F); 300-hPa winds (kt; 1 kt = 0.51 m s⁻¹) and height (m); 500-, 700-, and 850-hPa winds (kt), temperature (°C), and height (m); 500-hPa absolute vorticity (×10⁻⁵ s⁻¹); 850-hPa relative vorticity (×10⁻⁵ s⁻¹); sea level pressure (hPa), 10-m winds
(mi h\(^{-1}\)), and 3-h accumulated precipitation (in.); 0–72-h accumulated precipitation (in.); precipitable water (in.) and its standardized anomaly (e.g., Hart and Grumm 2001); 1000–500-hPa layer-mean relative humidity (%); surface-based CAPE (J kg\(^{-1}\)) and 10-m, 850-hPa, 700-hPa, and 500-hPa wind vectors (kt); and 850-hPa equivalent potential temperature (K). Such charts are provided in 3-h increments for the GFS and NAM data and 6-h increments for the ECMWF data. From the HPC operational forecasts, forecasters are provided with forecast charts of 0–72-h accumulated precipitation (in.) in 6-h increments. Additionally, forecasters are provided with a full suite of National Hurricane Center advisory products, valid at 0300 UTC 22 August 2008, in text format.

For the second stage of the forecasting exercise, ensemble forecast system forecast information is provided to forecasters primarily in the form of ensemble mean fields. In addition to the fields listed above, forecasters are provided with ensemble mean meteograms and charts of simulated composite reflectivity (dBZ). Forecasters are also provided with selected products designed to illuminate ensemble spread. These include the maximum and minimum 3- and 0–72-h accumulated precipitation from any ensemble member, the standard deviation of the 0–72-h accumulated precipitation forecast, matrix (or time series) charts of 3-h accumulated precipitation from each ensemble member, and plumes (e.g., http://www.spc.noaa.gov/expert/sref/fplumes/) of 0–72-h accumulated precipitation from each ensemble member. All fields are provided in 3-h increments. The units for each of these products are inches. The matrix and plume charts are provided for 15 locations in or near the Houston–Galveston CWA: New Iberia, Lafayette, and plume charts are provided for 15 locations in or near the Houston–Galveston CWA: New Iberia, Lafayette, and Waco, all located in Texas.

REFERENCES


