Changes in the Systematic Errors of Global Reforecasts due to an Evolving Data Assimilation System

THOMAS M. HAMILL

Physical Sciences Division, NOAA/Earth System Research Laboratory, Boulder, Colorado

(Manuscript received 13 March 2017, in final form 27 April 2017)

ABSTRACT

A global reforecast dataset was recently created for the National Centers for Environmental Prediction's Global Ensemble Forecast System (GEFS). This reforecast dataset consists of retrospective and real-time ensemble forecasts produced for the GEFS from 1985 to present day. An 11-member ensemble was produced once daily to +15-day lead time from 0000 UTC initial conditions. While the forecast model was stable during the production of this dataset, in 2011 and several times thereafter, there were significant changes to the forecast model that was used in the data assimilation system itself, as well as changes to the assimilation system and the observations that were assimilated. These changes resulted in substantial changes in the statistical characteristics of the reforecast dataset. Such changes make it challenging to uncritically use reforecasts for statistical postprocessing, which commonly assume that forecast error and bias are approximately consistent from one year to the next. Ensuring the consistency in the statistical characteristics of past and present initial conditions is desirable but can be in tension with the expectation that prediction centers upgrade their forecast systems rapidly.

1. Introduction

Statistical postprocessing refers to the adjustment of the current raw forecast guidance using statistical methods and time series of past forecasts and observations or analyses. Statistical postprocessing has a long heritage in many national weather services, decreasing systematic errors and improving probabilistic forecast skill and reliability (e.g., Glahn and Lowry 1972; Carter et al. 1989). In recent years, statistical postprocessing has increasingly been called upon to provide value-added guidance for difficult forecast problems, including high-impact weather such as heavy precipitation forecasts (Scheuerer and Hamill 2015, see their appendixes A–C) and for forecasts with lead times measured in weeks, not days (Hamill et al. 2004). In such situations, when a long time series of past forecasts (i.e., reforecasts) are available, they can be very useful in the postprocessing, helping distinguish the predictable signal amid the chaos-induced noise and the accumulating model bias. The author has twice now participated in the generation of global weather reforecast datasets, multidecadal retrospective forecasts using an operational forecast model (Hamill et al. 2006, 2013). The author has also worked with the U.S. National Weather Service to set up an infrastructure so that future reforecasts can be generated that are of high quality and statistical consistency.

This short manuscript describes a significant potential challenge with the production and use of reforecasts; namely, the challenges introduced when the reforecasts are initialized with a data assimilation system that is evolving. Even if the underlying forecast model is held fixed during a period of reforecast generation, there may be changes in systematic error characteristics of the underlying analysis due to changes in the data assimilation methodology, the type and number of observations, and the forecast model that provides its background. Consequently, the reforecast product may also have changes in its systematic errors, degrading their utility.

In the following sections, we will briefly describe the National Centers for Environmental Prediction (NCEP) Global Ensemble Forecast System (GEFS) reforecast dataset examined here and recent changes in the assimilation system (section 2). Section 3 provides some examples of changes in the bias of the reforecast system as a consequence of assimilation system changes.

Corresponding author: Thomas M. Hamill, tom.hamill@noaa.gov

DOI: 10.1175/MWR-D-17-0067.1

TABLE 1. A list of major implementation to the GFS and its data assimilation system that could have impacted the statistical characteristics of near-surface analyses.

Date	System	Change
9 May 2011	GFS	1) New thermal surface roughness length 2) Update to new version of Community Radiative Transfer Model
22 May 2012	GFS and GDAS	1) Incorporation of a hybrid 3D variational–ensemble data assimilation system 2) Various additional changes to the assimilation of remotely sensed data 3) Addition of new satellite wind data quality control
5 Sep 2012	GFS land surface	1) Change to a lookup table in land surface scheme that modulates evapotranspiration based on vegetation type and root zone depth to mitigate a warm-season cool bias in GFS
14 Jan 2015	GFS	1) Increase GFS horizontal resolution from T574 (~27 km) to T1534 (~13 km) 2) Change from Eulerian to semi-Lagrangian dynamics 3) Replace Reynolds SST analysis with 5-min daily real-time global SST 4) Initialize ice at small inland lakes in NH with 4-km multisensor snow and ice mapping system 5) Use 1982–2012 updated SST and sea ice climatology 6) Replace old update of model snow depth by direct insertion of U.S. Air Force–provided depth data with a blend of the model first guess and U.S. Air Force depth 7) Reduce drag coefficient at high wind speeds 8) Replace bucket soil moisture climatology with CFS/GLDAS climatology at T574 9) Replace 1° momentum roughness length climatology with lookup table based on vegetation type 10) Add a dependence of the ratio of thermal and momentum roughness on vegetation type 11) Increase ensemble Kalman filter resolution used in hybrid 3D variational data assimilation from T254L64 to T574L64 12) Assimilate hourly geostationary satellite atmospheric motion vectors 13) Use improved stochastic physics in EnKF ensemble forecasts
11 May 2016	GFS and GDAS	1) Change from 3D variational hybrid to 4D variational hybrid assimilation system 2) Addition of various satellite data (AMSU-A, AVHRR-derived winds) 3) Improve aircraft data bias correction 4) Community Radiative Transfer Model upgrade and bug fix 5) Correction of land surface characteristics for grassland and cropland to reduce summertime warm and dry biases over Great Plains 6) Improve convective gravity wave drag

Section 4 provides a discussion about the problems noted in this article and how they may be addressed when generating future reforecasts.

2. A description of the GEFS reforecast procedure and evolution of the underlying data assimilation system

Second-generation global ensemble reforecasts from NCEP GEFS were described more thoroughly in Hamill et al. (2013). We review salient details here.

Once-daily (from 0000 UTC initial conditions), 11-member reforecasts and real-time ensemble forecasts have been generated from 1 January 1985 to present day using the NCEP GEFS model system as was implemented operationally at 1200 UTC 14 February 2012. This was version 9.0.1 of the GEFS (discussed online at http://www.emc.ncep.noaa.gov/GFS/impl.php). There is a known bug in version 9.0.1 that resulted in the use of incorrect land surface tables in the land surface parameterization, and the effects of this bug contaminated the near-surface temperatures. This buggy model version was used for all forecasts for consistency.

During the period of the reforecast and real-time forecasts, the underlying data assimilation system has changed multiple times (Table 1). Through 20 February 2011, control initial conditions were generated by the Climate Forecast System Reanalysis (CFSR; Saha et al. 2010), computed with a three-dimensional (3D) variational data assimilation scheme using background forecasts from a specially designed version of the Global Forecast System (GFS) at T382L64 resolution (i.e., spectral triangular truncation at wavenumber 382 with 64 vertical levels). From 20 February 2011 to May 2012, initial conditions were taken from the operational Gridpoint Statistical Interpolation analysis system (GSI), with a somewhat different version of the GFS and T574L64 resolution. After 22 May 2012, the GSI was upgraded to use a hybrid ensemble Kalman filter/3D variational analysis system (Kleist and Ide 2015). This analysis improved the skill of operational GEFS forecasts and thus of the reforecasts introduced into the archive subsequent to that date. Several other significant implementations followed, including a correction to the land surface table bug fix in the underlying GFS system in September 2012 (though not in the GEFS system), a large number of changes to the GFS and GSI systems in January 2015, and a change from 3D to 4D hybrid ensemble–variational analysis in May 2016. More details on these changes are provided in Table 1. The main point here is that prior to the 2012 version of the GEFS becoming operational, the initial conditions were obtained from CFSR, and subsequent reanalyses and reforecasts were created. Thereafter, the real-time forecasts from GEFS version 9.0.1 were archived; though the GEFS forecast system was fixed, the underlying control analyses feeding them changed significantly.

3. Examples of temporal changes in the GEFS reforecast characteristics

In this section we focus on changes to the thermodynamic and precipitation characteristics of analyses and forecasts in the GEFS reforecasts. In particular, we will consider their characteristics in the central to eastern United States, where in recent years there have been notable biases in the GFS near-surface analyses (M. Ek 2016, personal communication). Consider cumulative distribution functions (CDFs) of convective available potential energy (CAPE; Bluestein 1993, his section 3.4.5), shown in Fig. 1. Again, prior to 2011, GEFS reforecasts were initialized from the CFS reanalysis, and afterward from the real-time analysis. The April–June CDFs of analyzed CAPE indicate that the frequency distribution subsequent to 2011 was shifted to dramatically higher CAPE relative to before 2011. For example, the 80th percentile of analyzed CAPE was $800\,\mathrm{J\,kg^{-1}}$ prior to 2011 and approximately $2100\,\mathrm{J\,kg^{-1}}$ thereafter. This analysis bias affected the forecasts as well, with shorter-range forecasts showing more of an effect of the analysis change than longer-lead forecasts (blue curves). By +120 h, regardless of the initial analysis, distributions of forecast CAPE were relatively similar, indicating that the GEFS forecasts adjusted to the intrinsic bias of that version of the prediction system used in the reforecast. The implications of this are that distributions of reforecast CAPE do not have anything close to stationary error statistics for forecasts with short lead times. Hence, reforecast-based postprocessing methodologies utilizing CAPE as a predictor must account for this change in character in order to provide meaningful results.

What underlies this change in CAPE, changes in temperature and/or changes in moisture analyses? This article will not examine changes above the surface, but

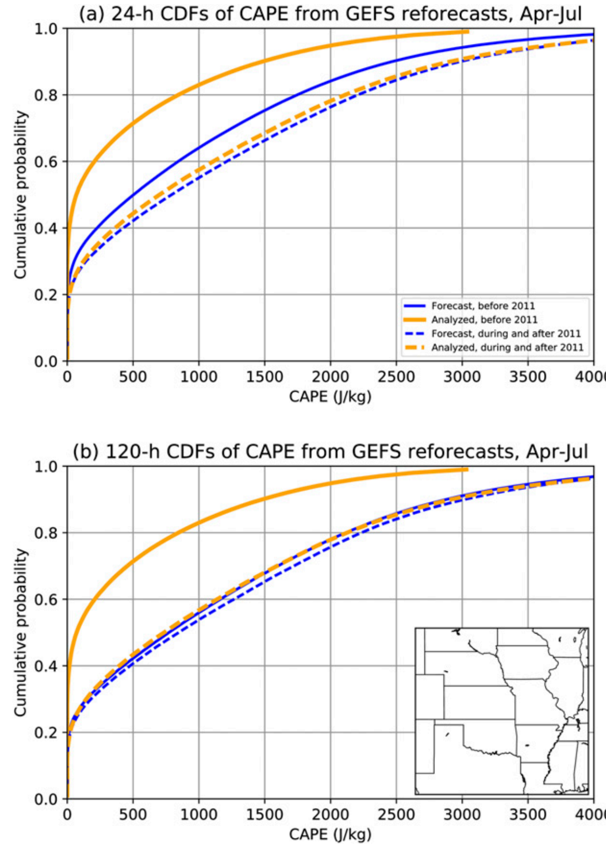

FIG. 1. Changes in the cumulative distribution function of analyses and forecasts of convectively available potential energy (CAPE) from GEFS reforecast initialization prior to vs during/after 2011. Data composited over April–June data for the region shown in the inset box. Red lines denote analyses, and blue lines denote reforecasts. Solid lines indicate data from before 2011, and dashed lines for data during and after 2011. Forecast CAPE at the (a) 24-h lead and (b) 120-h forecast lead. Inset box shows the domain where CDFs were calculated in the U.S. Great Plains.

Fig. 2 provides information on differences in temperature and dewpoint 2 m above the surface with respect to ERA-Interim (Dee et al. 2011), a reanalysis that used a stable forecast model and assimilation system. Figure 2a shows ERA-Interim temperature and dewpoints averaged over data from the same region shown in Fig. 1b. Notice the annual cycle of monthly mean temperature, with only modest departures from year to year, consistent with interannual variability. Figure 2b then shows the differences of the GEFS initial state in this region relative to ERA-Interim. The most noticeable difference is that subsequent to 2011, the GEFS 2-m differences from ERA-Interim indicate that the GEFS become markedly moister. Dewpoint differences (GEFS minus ERA-Interim) jump 1°–3°C in the warm season, relative to their differences prior to 2011. The

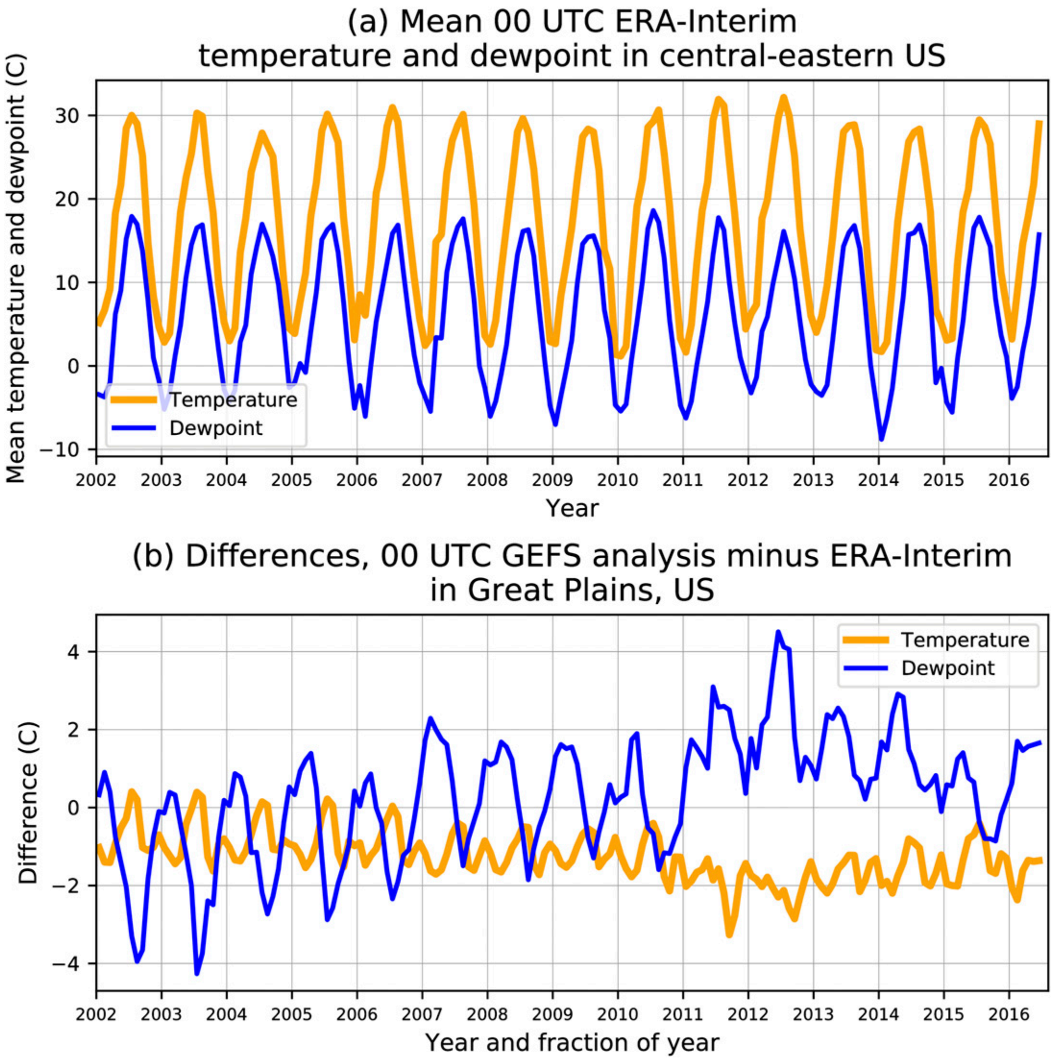

FIG. 2. (a) Time series of 2-m temperatures and dewpoints at 0000 UTC from ERA-Interim using the same box shown in Fig. 1b. (b) Time series of mean differences at 0000 UTC between the temperature of the GEFS initial analysis and the ERA-Interim analysis for temperature (orange curve) and dewpoint (blue curve).

0000 UTC temperature differences also change; GEFS 0000 UTC analyses become more markedly cool relative to ERA-Interim, especially in the 2011–14 time period. Since CAPE calculations are more sensitive to dewpoint perturbations, the increase in the analyzed moisture subsequent to 2011 are likely responsible for the increase in CAPE seen in Fig. 1.

Did the character of precipitation forecasts also change markedly subsequent to 2011? Since precipitation distributions are often well fit with modified Gamma distributions [Scheuerer and Hamill (2015) and references therein], we first consider the characteristic distributions fitted to forecast and analyzed data. Gamma distributions are used for the fits and represent average parameters over the same region shown in Fig. 1b. Rather than using the more involved censored, shifted Gamma distributions of Scheuerer and Hamill (2015), here we fit three parameters: (i) the percentage of samples with zero precipitation, and for the remaining samples with nonzero precipitation, the fitted Gamma distribution (ii) shape (α) and (iii) scale (β) parameters. Fitted parameters used the maximum likelihood estimator approach of Thom (1958) discussed in Wilks (2011). Data are shown for samples of GEFS reforecast and $1/8°$ Climatology-Calibrated Precipitation Analysis (CCPA) data (Hou et al. 2014). Figure 3 shows substantial annual and interannual variability of the fitted parameters, but it does not show any readily apparent systematic change subsequent before versus after 2011, as was seen with CAPE and dewpoint.

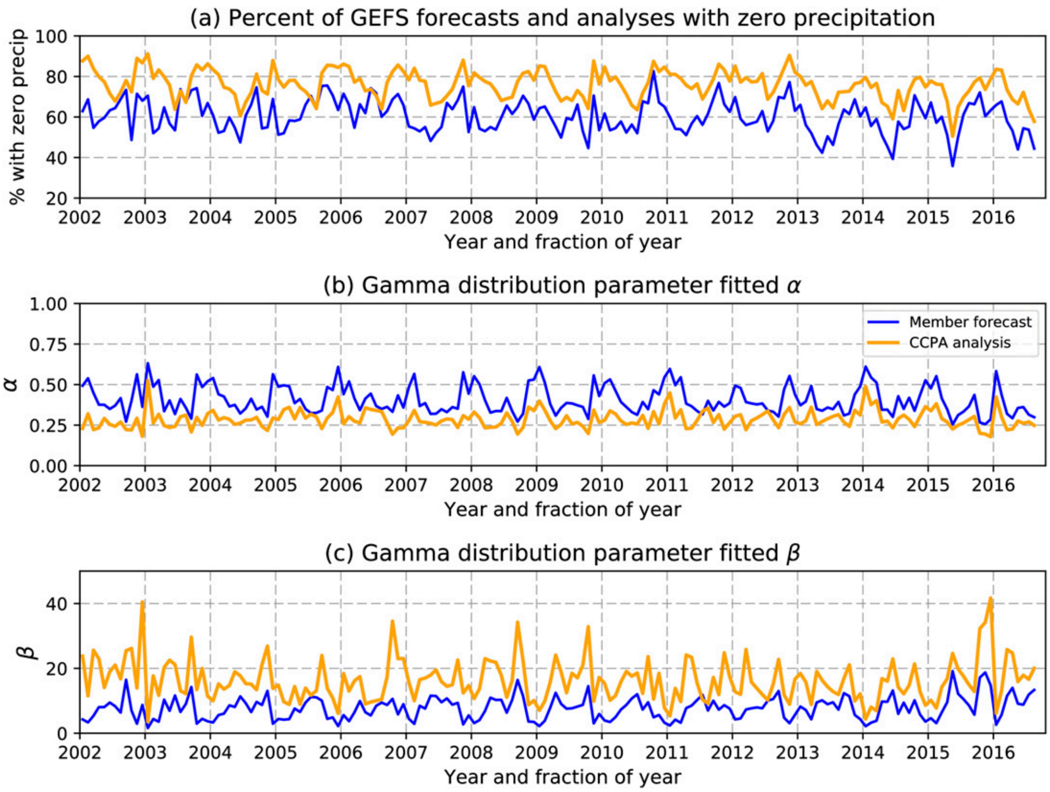

FIG. 3. Time series of parameter fits of precipitation for CCPA analyses (orange) and 12–24-h GEFS member forecasts (blue) in the region shown in Fig. 1b. Separate parameter fits were prepared for each month of the year from 2002 to 2016. (a) The percentage of grid points reporting zero precipitation, (b) the fitted shape parameter for analysis and forecast, and (c) the fitted scale parameter.

It is still possible that regression relationships between forecast and observed may have changed during that period. To examine this, we fit an extended logistic-regression model (Wilks 2009) to postprocess the precipitation data. This postprocessing method permits the estimation of a full probability distribution from the input data. For a given precipitation amount q, the probability of equaling or exceeding q is assumed to follow the functional form as

$$p(q) = \frac{\exp[\beta_0 + \beta_1 g(q) + \beta_2 \underline{x}]}{1 + \exp[\beta_0 + \beta_1 g(q) + \beta_2 \underline{x}]}, \quad (1)$$

where β_0, β_1, and β_2 are fitted parameters and $\underline{x}$ is the power-transformed ensemble-mean precipitation amount. Precipitation forecasts were transformed with a square root transformation (ibid.) and used the function $g(q) = \sqrt{q}$ (ibid.). In this approach, data were pooled across the geographic region of interest and fit using all data during the month of interest. Training was performed simultaneously over amount thresholds of 0.4, 1.0, 2.5, 5.0, and 10 mm. Figure 4 shows the time series of fitted extended logistic distribution parameters for +12- to +24-h forecasts. It does not appear that the fitted parameters after 2011 are statistically inconsistent with the fitted parameters before 2011, though there is some suggestion that the intercept parameter β_0 may differ before versus after 2011. Figure 5 illustrates $\geq$5 mm $(12\,\mathrm{h})^{-1}$ probability forecasts based on these extended logistic regression models as a function of the year/month and the forecast amount. From visual inspection, there does not appear to be a noticeable change in the regression model before versus after 2011. This suggests that precipitation forecast data may not be as strongly affected as for the thermodynamic information, somewhat surprisingly.

4. Discussion and conclusions

The challenges of generating a reforecast with stable forecast-error characteristics was demonstrated in this article. Even if reforecasts were generated using the same forecast model as used in the real-time system, should the analyses used in the forecast initialization change, then the characteristics of the reforecasts can change as well. In the example shown here, GEFS reforecasts prior to 2011 used data from the CFS

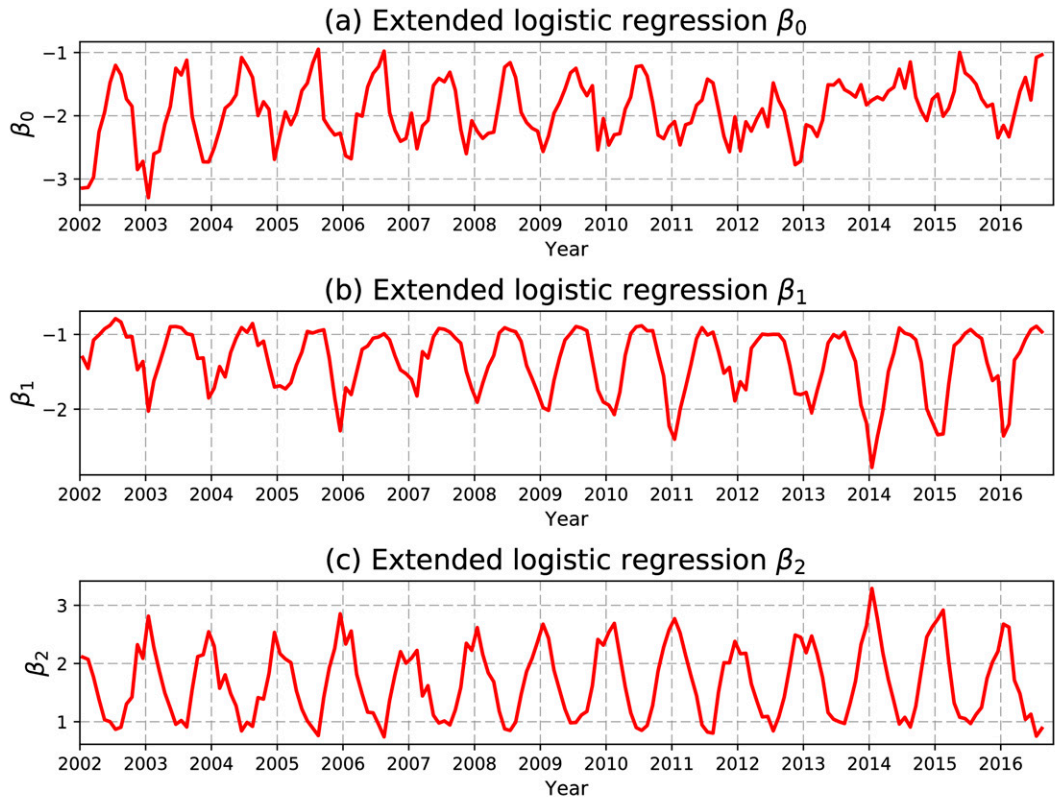

FIG. 4. Time series of monthly fitted extended logistic regression parameters using GEFS ensemble-mean reforecasts trained against 1/8° CCPA data for points in the box shown in Fig. 1b.

reanalysis and thereafter used real-time analyses. These changed several times in 2011 and thereafter. The effects were particularly noticeable in short-range forecasts of thermodynamic variables. One might expect that other prediction centers that use reforecasts such as the European Centre for Medium-Range Weather Forecasts (ECMWF) might also have similar problems with their reforecasts, some of which are currently initialized from ERA-Interim (Dee et al. 2011). This article did not examine ECMWF data, however.

While reforecasts are strongly desired for many applications, including precipitation postprocessing, hydrologic forecast system validation, and the postprocessing of longer-lead forecasts, it is apparent that thought and care must be put into how a reforecast system is configured. Suppose an ensemble reanalysis and reforecast are generated with the current assimilation and forecast system (currently in the U.S. National Weather Service, these are based on hydrostatic spectral global models). Thereafter, the deterministic forecast

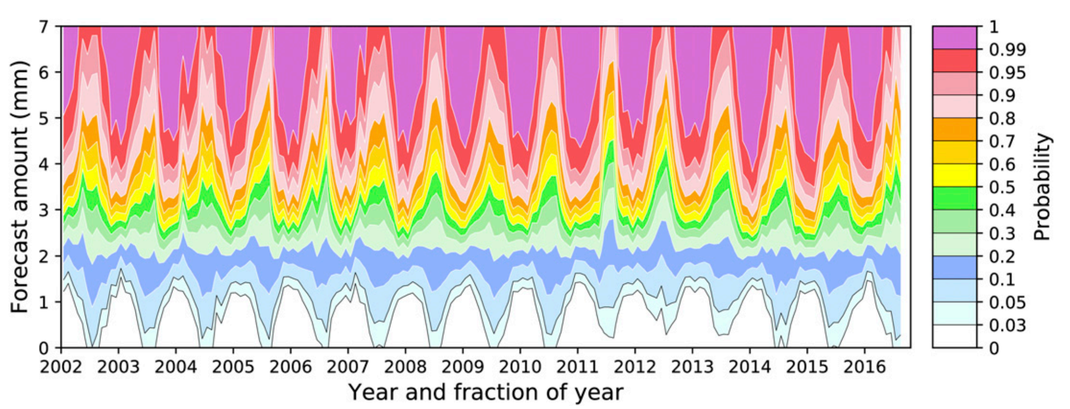

FIG. 5. Probability forecasts based on regression model using parameters from Fig. 4.

model and the forecast model used in the data assimilation system changes, perhaps to a new dynamical core, as indeed the United States anticipates doing in the next few years. In such a case, the statistical characteristics of the (spectral based) reanalysis differ greatly from the characteristics of the eventual (grid point based) real-time analyses. The reforecast will inherit such differences, making the dataset nonstationary and more difficult to use in postprocessing.

While reanalysis and reforecasting may be necessary to provide the long training and validation datasets needed for many applications, they must be constructed carefully. Such datasets are very computationally expensive to generate. To provide a sufficient return on such an investment, some guiding principles for their construction are proposed. 1) If major system changes are anticipated, it is preferable to generate a new reanalysis and reforecast after the system has changed and proven stable rather than before. Should this advice be ignored, it may become apparent that a new reanalysis and reforecast are necessary a few scant months or years after the last one starts to be used operationally. 2) Sometimes major changes to a forecast system are necessary. Arguably, a change to a new dynamical core that permits the explicit prediction of thunderstorms is one of those necessary changes. Some other changes, however, might provide only slight improvements to the RMS errors of forecasts but might notably change the bias characteristics. The possible effects on previously generated reforecasts might thus be a new and useful criterion to evaluate when deciding whether a proposed change is implemented. 3) Related to 2), it may be preferable to build systems that maintain a low and consistent bias, even if RMS errors may be higher than what is possible with a more biased system. When we consider reforecasts and their use in postprocessing as part of the system, then small improvements in error accompanied by large changes in bias may degrade rather than improve the final product, unless new reanalyses and reforecasts can be generated again. Perhaps this may motivate operational prediction centers to attempt to bias correct the background forecasts in the data assimilation (Dee 2005).

National weather services are increasingly embracing the regular generation of reanalyses and reforecasts, as they can tremendously improve the skill of the final numerical guidance via postprocessing. However, these technologies cannot simply be unthinkingly bolted onto an existing prediction system. Seeing numerical weather prediction as a holistic process including postprocessing, we should change our procedures for evaluating potential changes to our prediction system; postprocessed skill and stability of biases become criteria to consider as well as raw numerical skill.

Acknowledgments. This work was supported by funding provided to NOAA/ESRL/PSD by NOAA/NWS/STI under the Next-Generation Global Prediction System Grant P8MWQNG-PTR.

REFERENCES

Bluestein, H. B., 1993: *Synoptic-Dynamic Meteorology in Midlatitudes.* Vol. II, *Observations and Theory of Weather Systems*, Oxford Press, 594 pp.

Carter, G. M., J. P. Dallavalle, and H. R. Glahn, 1989: Statistical forecasts based on the National Meteorological Center's numerical weather prediction system. *Wea. Forecasting*, **4**, 401–412, doi:10.1175/1520-0434(1989)004<0401:SFBOTN>2.0.CO;2.

Dee, D. P., 2005: Bias and data assimilation. *Quart. J. Roy. Meteor. Soc.*, **131**, 3323–3343, doi:10.1256/qj.05.137.

——, and Coauthors, 2011: The ERA-Interim reanalysis: Configuration and performance of the data assimilation system. *Quart. J. Roy. Meteor. Soc.*, **137**, 553–597, doi:10.1002/qj.828.

Glahn, H. R., and D. A. Lowry, 1972: The use of model output statistics (MOS) in objective weather forecasting. *J. Appl. Meteor.*, **11**, 1203–1211, doi:10.1175/1520-0450(1972)011<1203:TUOMOS>2.0.CO;2.

Hamill, T. M., J. S. Whitaker, and X. Wei, 2004: Ensemble reforecasting: Improving medium-range forecast skill using retrospective forecasts. *Mon. Wea. Rev.*, **132**, 1434–1447, doi:10.1175/1520-0493(2004)132<1434:ERIMFS>2.0.CO;2.

——, ——, and S. L. Mullen, 2006: Reforecasts, an important dataset for improving weather predictions. *Bull. Amer. Meteor. Soc.*, **87**, 33–46, doi:10.1175/BAMS-87-1-33.

——, G. T. Bates, J. S. Whitaker, D. R. Murray, M. Fiorino, T. J. Galarneau Jr., Y. Zhu, and W. Lapenta, 2013: NOAA's second-generation global medium-range ensemble reforecast dataset. *Bull. Amer. Meteor. Soc.*, **94**, 1553–1565, doi:10.1175/BAMS-D-12-00014.1.

Hou, D., and Coauthors, 2014: Climatology-calibrated precipitation analysis at fine scales: Statistical adjustment of stage IV toward CPC gauge-based analysis. *J. Hydrometeor.*, **15**, 2542–2557, doi:10.1175/JHM-D-11-0140.1.

Kleist, D. T., and K. Ide, 2015: An OSSE-based evaluation of hybrid variational–ensemble data assimilation for the NCEP GFS. Part I: System description and 3D-hybrid results. *Mon. Wea. Rev.*, **143**, 433–451, doi:10.1175/MWR-D-13-00351.1.

Saha, S., and Coauthors, 2010: The NCEP Climate Forecast System Reanalysis. *Bull. Amer. Meteor. Soc.*, **91**, 1015–1057, doi:10.1175/2010BAMS3001.1.

Scheuerer, M., and T. M. Hamill, 2015: Statistical postprocessing of ensemble precipitation forecasts by fitting censored, shifted Gamma distributions. *Mon. Wea. Rev.*, **143**, 4578–4596, doi:10.1175/MWR-D-15-0061.1.

Thom, H. C. S., 1958: A note on the Gamma distribution. *Mon. Wea. Rev.*, **86**, 117–122, doi:10.1175/1520-0493(1958)086<0117:ANOTGD>2.0.CO;2.

Wilks, D. S., 2009: Extending logistic regression to provide full-probability-distribution MOS forecasts. *Meteor. Appl.*, **16**, 361–368, doi:10.1002/met.134.

——, 2011: *Statistical Methods in the Atmospheric Sciences.* 3rd ed. Academic Press, 676 pp.