A New Measure of Ensemble Central Tendency

Jing Zhang[1], Jie Feng[2]

[1]*Global Systems Division, ESRL/OAR/NOAA, Boulder, CO, USA*

[2] *School of Meteorology, University of Oklahoma, Norman, OK, USA*

Final Project Report to Developmental Testbed Center (DTC) Visitor Program, 2018

1. Introduction

Due to the chaotic nature of the atmosphere (e.g., Lorenz 1963, Yuan et al. 2018), errors in Numerical Weather Prediction (NWP) originating from the use of imperfect initial conditions and numerical models inevitably amplify. In addition to producing a single unperturbed or control forecast from the best available initial condition, a properly formulated ensemble of forecasts may also offer some value (e.g., Toth and Kalnay 1993; Molteni et al. 1996). One benefit is that the mean of an ensemble (generally defined as the Arithmetic Mean (AM) of ensemble member forecasts) filters out features that are out of phase in the member forecasts. These, typically finer scale features have little or no skill, hence AM is characterized with a root-mean-square error (RMSE) lower than that in a single unperturbed forecast (e.g. Toth and Kalnay 1997). Consequently, in the past two decades ensemble mean forecasts became widely used and important products at operational forecast centers across the globe.

Notwithstanding its popularity and value, the use of AM as an ensemble central tendency has its limitations. First, as pointed out for example by Molteni et al. 1996, Toth and Kalnay 1997, and Surcel et al. 2014, the filtering out of phase features in AM results in reduced variability. In particular, the amplitude of features that are misaligned across ensemble members is reduced in AM. In other words, the elimination of less- or unpredictable features makes AM unrealistically smooth both in space and time. For example, AM renders the sharp low-pressure wave present in all ensemble members at various longitudes as a wide and shallow low-pressure system (see solid blue line in

Fig. 1(a)). Consequently, the Cumulative Distribution Function (CDF) of AM forecast values over a domain is also changed, eliminating extremes and reducing the range of values as compared to individual analysis or forecast fields (Ebert 2001). Such suboptimal CDF results in a loss of dynamical and physical consistency and spatial covariance present in the individual forecast fields and across different variables and levels in the conglomerate of features present in AM. For these reasons, AM fields may be confusing or misleading, and are notoriously challenging to use (Ebert 2001; Knutti et al. 2010; Feng et al. 2019). The disadvantages of AM in many meteorological applications stem from its pointwise, univariate definition: $\bar{x}_k = (1/N) \sum_{i=1}^{N} x_{i,k}$, where $x_{i,k}$ is a single variable at grid point k of the ith ensemble member (totally N members). It is apparent that strong spatial and temporal covariances present in AM are ignored.

Various methods have been proposed to alleviate the disadvantages of AM. The ensemble median was introduced as an alternative to AM (Galmarini et al. 2004; Zhou and Du 2010). While in the presence of outlier member forecasts the median may have some advantages compared to AM (Delle Monache et al. 2006), they are statistically nearly identical otherwise. To restore the CDF that the AM procedure distorts, Ebert (2001) suggested to relabel the contours in AM in such a way that the CDF in the proposed probability-matched mean exactly matches that in the constituent ensemble member forecasts. This manipulation, however, restores only the CDF; it will not undo the distortion, smoothing, and somewhat arbitrary positioning of the forecast features.

In addition, since spatially coherent features among ensemble forecast samples contain displacement (or position) and amplitude errors (Hoffman et al. 1995), in recent papers, Ravela (coalescence, 2012, 2013, 2014) and Purser (generalized ensemble mean or GEM, 2013) considered the generalization of central tendency for covarying multivariate variables. Using covariances estimated from the ensemble itself, coalescence deforms features in each ensemble field to their mean position using the Field Alignment (FA) technique of Ravela (2007a) so that given some general constraints, the difference between individual displaced ensemble fields and their mean are minimized. In the iterative variational minimization both the mean field and the deformation for each member are estimated. GEM is constructed similarly except for a slightly different variational minimization algorithm.

Note that both methods require knowledge of the covariance between forecast fields (see Eq. 8 of Ravela 2012 and Eq. 1.5 of Purser 2013), estimation of which remains a challenging problem despite decades of efforts in ensemble-based data assimilation (Hamill et al. 2001; Houtekamer et al. 2005; Wang et al. 2013). When applied to calculate coalesced or generalized ensemble mean fields, for practical application to operational ensemble forecasts, the objective would become more complex with higher model resolution and more ensemble members, apparently increasing the computational expense.

In this paper we introduce a simplified and computationally efficient method for the direct, vector-based calculation of generalized mean fields (Section 2). The non-

variational method called Feature-oriented Mean (FM) requires no use of explicit covariance information; instead, it exploits such information implicit in the ensemble forecasts. FM will be tested and evaluated using operational ensemble forecasts (Section 3). Evaluation of results including a comparison with AM will be shown in section 4 while a summary and discussion are offered in Section 5.

2. Methodology
a. Field Alignment

Atmospheric motions manifest in spatiotemporally coherent features. Features can be characterized, for example, by their geographical position, amplitude, or other characteristics (Hoffman et al. 1995; Ravela et al. 2007a; Beezley and Mandel 2008). Notwithstanding, a large body of traditional meteorological research and operational applications use an observational, or gridded point-wise approach when comparing states of the atmosphere, disregarding spatiotemporally organized structures. In general, differences between atmospheric states (e.g., forecast and verifying analysis – 2D forecast error, or unperturbed and perturbed ensemble forecasts – 2D perturbation fields) can be, with some assumptions decomposed into a positional and a residual (or amplitude) error (e.g., Hoffman et al. 1995; Ravela et al. 2007a; Peña et al. 2019).

The FA technique is designed for such a decomposition of forecast error or difference fields. For easy access by the community, FA was ported into the Developmental Testbed Center (DTC) Code Repository in a previous study funded by the DTC Visit Program (PI: Said Ravela).

If **Y** is a 2D field similar to **X** except its features are somewhat displaced, FA defines, and variationally estimates a smooth 2D displacement vector field $\vec{D}$ that if applied as a translation operation to each point of **Y** (**Y** adjusted to **Y'**), will minimize the remaining (henceforth called amplitude) RMS difference between **X** and **Y'**. The smoothing of the displacement vector is conducted by a spectrum truncation algorithm (Ravela 2012). It has a unique tunable wavenumber parameter l to determine those larger scale spatial coherent "features" to be transposed, such as a collection of cold front cases, while those finer scale structures that are less predictable or unpredictable will not be particularly adjusted but move along with the larger scales. Unlike other error decomposition techniques developed by Hoffman et al. (1995), Du et al. (2000), and Nehrkorn et al. (2003, 2014), FA uses less tunable parameters (i.e. the unique smoothing parameter l) and does not rely on the posterior (i.e. after alignment) forecast error covariance information. FA has been applied in a wide range of application areas including data assimilation (Ravela 2007b), verification (Ravela 2007a, 2014; Peña et al. 2019), nowcasting (Ravela 2012), and spatiotemporal error propagation (Feng et al. 2017).

b. Feature-oriented Mean

The FA technique provides a fundamental tool for the FM scheme. The key concept of FM is to align the features of fields with spatial coherence to the mean position. The finer scales that are unpredictable are not considered features but rather they are termed noises. Assume that x_j is a randomly selected member of an N-member

ensemble x_i ($i=1, 2, 3, \ldots, N$). The working definition of FM consists of the following five steps:

(1) Compute the displacement vector $\overrightarrow{D_{ji}}$ between x_j and each of the other N-1 members x_i by using the FA technique.

(2) Calculate the average of displacement vectors $\overrightarrow{D_j} = (1/N)\sum_{i=1}^{N}\overrightarrow{D_{ji}}$. $\overrightarrow{D_j}$ represents the displacement of features in x_j compared to the mean of the position of features in all ensemble members.

(3) Adjust member x_j by transposing its 2D field in space by the displacement vector $\overrightarrow{D_j}$: $x_j^{'} = x_j + \overrightarrow{D_j}$. This adjustment will align the position of features in member x_j to the mean of their position in the entire ensemble.

(4) Repeat steps (1) - (3) for each member of the ensemble. The aligned members (see green lines in Fig. 1(b)) will differ only in the amplitude of their features, along with the incoherent small-scale noises unaffected by the FA procedure.

(5) Feature-oriented Mean (FM) is defined as the arithmetic mean of all aligned ensemble members: $\bar{x} = (1/N)\sum_{i=1}^{N}x_i^{'}$ (see red solid line in Fig. 1(c)).

3. Experimental setup

The FM algorithm described in section 2.b is evaluated and compared to AM using forecasts from the National Center for Environmental Prediction (NCEP) Global Ensemble Forecasting System (GEFS, Toth and Kalnay 1993, 1997; Zhu et al. 2012; Zhou et al. 2016, 2017). 20-member 500 hPa Geopotential Height (GH) 00Z GEFS forecasts from a 25-day sample period (01 Oct – 25 Oct 2013) are compared to

corresponding verifying analysis fields from the Global Forecast System (GFS) on a common 1-degree horizontal resolution grid. Ensemble forecasts offer an ideal testing ground for the use of FM as the larger scale; more predictable features possess a coherence among ensemble members that FM can readily detect. The smoothing parameter l is set to be 128 which is recommended by Ravela, the author of the FA code (personal communication).

An important comparison will be the level of smoothing imposed by AM vs. FM, as revealed by a comparison of the ensemble mean energy spectrum with those of analyses. A better forecast should have more consistent and less smoothed power spectrum compared to analyses. The overall forecast skill will also be evaluated by traditional metrics like Pattern Anomaly Correlation (PAC) and RMSE. With these metrics the performance of AM and FM will be compared for a selected case and sample mean results.

4. Results
a. A Case Study

The transposition of features in an individual ensemble member to the mean position of features in all members for a case of a 7-day Northern Hemisphere (NH, 20°-80°) forecast initialized at 0000 UTC 12 Oct is shown in Fig. 3. The original forecast (blue) is transposed (red, step 3 in FM algorithm, cf. Fig 2) with the displacement vector field (black arrows, step 2). As seen from the small displacement vectors and correspondingly small displacement of contour lines over the Atlantic,

features in the selected ensemble member (member 10) align well with the mean position of features in the rest of the ensemble members in this geographic area. On the contrary, the selected ensemble member (member 10) appears to be an outlier over the Pacific, as manifested by the large displacement vectors and correspondingly large displacement of contour lines around 160W. This is confirmed by Fig 4a where the heavy copper color spaghetti line for member 10 is seen as an outlier among the other members.

As seen in Fig 4b, the FM algorithm aligns the features in member 10, along with those in the rest of the members, with the mean of the features' position in the original members. As seen in Fig. 4c, this results in an FM field that when compared with AM, better reflects the consensus in the position, and especially in the amplitude of features in the original ensemble. In other words, FM better preserves the consensus in the features among the averaged fields, a potential advantage in either synoptic forecast or climatological applications.

b. Amplitude as a function of lead time and scales

In Fig. 5 we quantify how much more total amplitude (defined against the climatological mean) FM retains over AM as a function of lead time, averaged over the global domain and all 25 cases. Results are stratified according to the AM anomaly forecast at any point in time and space being below 1, between 1-2, and above 2 climatological standard deviations. For all categories, FM retains up to 14% more amplitude compared to AM.

The spectral distribution of analysis and various lead time AM and FM forecast anomalies is displayed in Fig. 6. As perturbation evolution at the early 2-day lead time is mostly linear (Gilmour et al. 2001), only minor differences are seen between the two types of mean forecasts and only at the finest scales. As with increasing lead time nonlinearities emerge on larger scales, the variance in AM falls below that of natural variability in the analysis at progressively larger scales. FM, meanwhile, retains more natural variability on those scales, due to the alignment of features before the mean of the member forecasts is taken.

c. Error metrics

It is well understood that the RMSE in sample-based statistical estimates of the expected value of a quantity is minimized by AM (Li et al. 2018; Feng et al. 2019) due to its (sample size dependent) reduction of the variance (i.e., noise) in the sample. Any deviation from the AM formula in Eq 1 can only increase RMSE. It is evident that the retention of more variance in FM increases RMSE (Fig. 7a) and reduces the positively oriented forecast performance measure of Pattern Anomaly Correlation (PAC, Fig.7b) only slightly (cf. solid black and red curves) for NH. The results for SH are similar (not shown). As expected, Fig. 7 also shows that the skill of the members whose features are aligned with the mean of the position of features in the original members (dotted red) is much better than that of the original members (dotted blue), with a much narrower range of variability in skill (cf. red vs. blue vertical bars). This is because the alignment eliminates much of the position related error introduced by the addition of initial

ensemble perturbations. A related study (Peña et al 2019) explores the decomposition of forecast errors into positional and amplitude/structural errors in more depth.

Interestingly, we find that for gridpoints over NH with large (i.e., larger than 1.5 standard deviation) observed (based on analysis data) anomalies, the error in FM forecasts, on average, is up to 10% lower than that in AM, with favorable FM performance in up to 70 % of the cases at about day 7 (Fig. 8). The results for SH are similar (not shown). What is behind this behavior and whether and how it could be exploited in prognostic applications remain to be explored in future studies.

5. Conclusions and discussion

Arithmetic mean (AM), when applied to an ensemble, reduces forecast error by filtering out part of the unpredictable variance present in individual members (i.e., waves completely out of phase). For this reason, AM gained widespread use in weather forecasting. On the other hand, when applied in a traditional, univariate sense, AM may reduce the amplitude, and distort the structure of partly predictable features that are present at somewhat different locations and with somewhat varying structures in the perturbed ensemble forecasts. Recognizing the spatiotemporal coherence of partly predictable features across ensemble members, we propose to spatially co-locate such features before their mean is taken. In the new ensemble central tendency that we call feature-oriented mean (FM), all forecast features appear at the mean of their position in the individual members, represented with an amplitude that is the mean amplitude of features aligned in all members to their mean position. Instead of averaging a collection

of assumedly uncorrelated variables like in AM, FM estimates the expected state of a multivariate system, composed of features with covarying elements.

The concept of central tendency generalized for multivariate systems with co-varying variables has been previously suggested by Ravela (2012, Coalescence) and Purser (2013, generalized mean, GM). The algorithmic implementation of the concept in these studies, however, is computationally expensive while meteorological applications are lacking. In the current study, we present a more efficient and readily parallelizable algorithm with an application to ensemble averaging. Both FM and Ravela's (2012) Coalescence method (a) calculate the mean amplitude field by (b) aligning features in each ensemble field to the mean of their position in the individual fields, using (c) the field alignment (FA) technique of Ravela (2007a) as a core technique. Coalescence, however, solves a more complex variational minimization problem to estimate the (i) mean amplitude field and (ii) displacement fields for each member at once, while FM solves a set of similar FA minimization problems followed by a simplified vector calculation to derive the estimation of mean position, after which it transposes each member with their corresponding displacement vector field. Note that unlike FM, both Coalescence and the GM algorithm assume the use of covariance information about the system being estimated which is usually difficult to, and poorly estimated (Hamill et al. 2001; Houtekamer et al. 2005; Wang et al. 2013), making their use more problematic.

On the computational side, the bulk of the FM algorithm (Steps 1-4 in Section 2.b) pertains to single ensemble members, perfectly suited for parallel processing on individual single cores. A further speed-up can be achieved if data are processed sequentially by lead time where the displacement vector solution from the previous lead time can be used as a first guess in the FA calculations.

Results from preliminary tests using operational NWP forecasts from the NCEP GEFS indicate that by the alignment of coherent larger scale, partly predictable features, FM retains up to 14% more variability than AM, particularly on partly and less predictable scales, resulting in features with more realistic amplitude. Meanwhile, forecast performance is not compromised as FM RMSE and PAC is hardly changed compared to AM. Interestingly, FM outperforms AM under extreme observed conditions, particularly for medium-range forecasts (e.g., 3-8 days lead time for synoptic scale). At earlier lead times characterized with quasi-linear evolution of ensemble perturbations, FM and AM have small differences, while at extended ranges where nonlinear saturation has a strong influence on the evolution of perturbations (e.g., 10 days and beyond for synoptic scales) FM may no longer be applicable. The latter is because at these longer lead times the features in different members become less similar, potentially preventing the variational FA algorithm in step (1) of the FM algorithm to converge to a displacement vector field solution. As the prediction of extreme events is more influential and critical than normal ones, and remains as a weakness of AM, FM

offers a potentially more useful tool for predicting such type of events than AM. A potential benefit needs to be further explored in future studies.

For simplicity, the smoothing coefficient in the reported FA experiments was fixed over all lead times. Ideally, one would allow the smoothing parameter l to decrease as a function of lead time, allowing the displacement vector field to initially contain most of the scales, while later only coarser scales, reflecting the upscale propagation of spatial scales dominating quasi-linear error growth. At long lead times when all predictability is lost and ensemble members constitute a random draw from climatology (i.e., no coherency in structures across ensemble members), smoothing can approach its maximum level (i.e., only a single, or no spatial adjustment over the entire domain), so FM asymptotes to AM.

Though FM here was demonstrated with 500 hPa height data, FM can be applied to any variable of interest, including non-continuous variables such as precipitation. Besides dynamically generated ensembles like the GEFS, FM can also be used to derive a consensus forecast from any set of NWP or other products like the multimodel superensemble forecasting (Ebert 2001; Krishnamurti et al. 2016). FM is expected to provide the most likely position and amplitude of severe weather events such as the evolution of a tropical storm prior to landfall, or a frontal zone with precipitation approaching a metropolitan area. Looking beyond weather forecasting, FM may be applicable in a wide array of synoptic and other type of climatology studies where a sample-based estimate of the typical behavior of various phenomena such as landfalling

hurricanes is sought. Once cases with the targeted phenomena present anywhere over a common domain are selected, with an appropriate level of smoothing, FM can be used to co-locate and then average the feature of interest.

Acknowledgements

REFERENCES

Beezley, J. D. and J. Mandel, 2008: Morphing ensemble Kalman filters. *Tellus* **60A**, 131-140.

Delle Monache, L., X. Deng, Y. Zhou, and R. Stull, 2006: Ozone ensemble forecasts: 1. A new ensemble design, *J. Geophys. Res.*, **111**, D05307.

Du, J., S. L. Mullen, and F. Sanders, 2000: Removal of distortion error from an ensemble forecast. *Mon. Wea. Rev*, **128,** 3347–3351.

Ebert, E. E., 2001: Ability of a poor man's ensemble to predict the probability and distribution of precipitation. *Mon. Wea. Rev.*, **129**, 2461-2480.

Feng, J., Z. Toth, and M. Peña, 2017: Spatially extended estimates of analysis and short-range forecast error variances. Tellus, 69A, 1325301.

Feng, J., J. Li, J. Zhang et al., 2019: The Relationship between deterministic and ensemble mean forecast errors revealed by global and local attractor radii. *Adv. Atmos. Sci.*, **36**, 271-278.

Galmarini, S., et al., 2004: Ensemble dispersion forecasting—Part I: Concept, approach and indicators, *Atmos. Environ.*, **38**, 4607–4617.

Gilmour, I., Smith, L. A., and Buizza, R. (2001) Linear regime duration: Is 24 hours a long time in synoptic weather forecasting? *J. Atmos. Sci.*, **58**, 3525–3539.

Hamill, T. M., J. S. Whitaker, and C. Snyder, 2001: Distance-dependent filtering of background error covariance estimates in an ensemble Kalman filter. *Mon. Wea. Rev.*, **129**, 2776–2790.

Hoffman, R. N., Z. Liu, J. Louis, C. Grassotti, 1995: Distortion representation of forecast errors. *Mon. Weather Rev.*, **123**, 2758-2770.

Houtekamer, P. L., H. L. Mitchell, G. Pellerin, M. Buehner, M. Charron, L. Spacek, and B. Hansen, 2005: Atmospheric data assimilation with an ensemble Kalman filter: Results with real observations. *Mon. Wea. Rev.*, **133**, 604–620.

Knutti, R., R. Furrer, C. Tebaldi, and J. Cermak, 2010: Challenges in combining projections from multiple climate models. *J. Climate*, **23**, 2739–2758.

Krishnamurti, T. N., V. Kumar, A. Simon, A. Bhardwaj, T. Ghosh, and R. Ross, 2016: A review of multimodel superensemble forecasting for weather, seasonal climate, and hurricanes. *Rev. Geophys.,* **54**, 336–377.

Li, J. P., J. Feng, and R. Q. Ding, 2018: Attractor radius and global attractor radius and their application to the quantification of predictability limits. *Climate Dyn.*, **51**, 2359–2374.

Lorenz, E. N., 1963: Deterministic nonperiodic flow. *J. Atmos. Sci.*, **20**, 130–141.

Molteni, F., R. Buizza, T. N. Palmer, and T. Petroliagis, 1996: The new ECMWF Ensemble Prediction System: Methodology and validation. *Quart. J. Roy. Meteor. Soc.*, **122**, 73–119.

Nehrkorn, T., R. Hoffman, C. Grassotti, and J-F. Louis, 2003: Feature calibration and alignment to represent model forecast errors: Empirical regularization. *Quart. J. Roy. Meteor. Soc.*, **129,** 195–218.

Nehrkorn, T., B. Woods, T. Auligné, and R. N. Hoffman, 2014: Application of feature calibration and alignment to high-resolution analysis: Examples using observations sensitive to cloud and water vapor. *Mon. Wea. Rev.*, **142**, 686–702.

Peña, M., I. Jankov, S. Gregory, S. Ravela, Z. Toth, 2019: Decomposition of forecast errors using a field alignment technique. Manuscript, available from the authors.

Purser, R. J., 2013: Can we make a better ensemble average? Manuscript, available from the authors.

Ravela S., 2007a: Two new directions in data assimilation by field alignment. *Lecture notes in Computer Science, Proc. ICCS*, 4487:1147-1154.

Ravela, S., K. Emanel, D. McLaughlin, 2007b: Data assimilation by field alignment. *Physica D: Nonlinear Phenomena*, **230**, 127-145.

Ravela, S., 2012: Quantifying Uncertainty for Coherent Structures, *Procedia Computer Science*, 9:1187-1196.

Ravela, S., 2014: Spatial inference for coherent geophysical fluids by appearance and geometry. In: IEEE Winter Conference on Applications of Computer Vision, ISSN 1550-5790, p. 925–932, DOI: 10.1109/WACV.2014.6836005

Ravela, S., 2013: Generating a forecast by field coalescence, US Patent Application US 13/904, 058, [available at https://patents.google.com/patent/US9046627B2/en?inventor=ravela&oq=ravela]

Surcel, M., I. Zawadzki, and M. K. Yau, 2014: On the filtering properties of ensemble averaging for storm-scale precipitation forecasts. *Mon. Wea. Rev.*, **142**, 1093–1105.

Toth, Z., and E. Kalnay, 1993: Ensemble forecasting at NMC: The generation of perturbations. *Bull. Amer. Meteor. Soc.*, **74**, 2317–2330.

Toth, Z., and E. Kalnay, 1997: Ensemble forecasting at NCEP and the breed- ing method. *Mon. Wea. Rev.,* **125,** 3297–3319.

Wang, X., D. Parrish, D. Kleist, and J. Whitaker, 2013: GSI 3DVar- based ensemble-variational hybrid data assimilation for NCEP global forecast system: Single-resolution experiments. *Mon. Wea. Rev.*, **141**, 4098–4117.

Yuan, H., Z. Toth, M. Peña, and E. Kalnay, 2018: Overview of Weather and Climate Systems. Pp. 35, in: Handbook of Hydrometeorological Ensemble Forecasting, Eds: Qingyun Duan, Florian Pappenberger, Jutta Thielen, Andy Wood, Hannah L. Cloke and John C. Schaake. Springer, doi:10.1007/978-3-642-40457-3_10-1

Zhou, B., and J. Du, 2010: Fog prediction from a multimodel mesoscale ensemble prediction system. Wea. Forecasting, 25, 303–322.

Zhou, X., Y. Zhu, D. Hou, and D. Kleist, 2016: A comparison of perturbations from an ensemble transform and an ensemble Kalman filter for the NCEP Global Ensemble Forecast System. *Wea. Forecasting*, **31**, 2057–2074.

Zhou, X., Y. Zhu, D. Hou, Y. Luo, J. Peng, and R. Wobus, 2017: Performance of the NCEP Global Ensemble Forecast System in a parallel experiment. *Wea. Forecasting*, **32**, 1989–2004.

Zhu, Y., D. Hou, M. Wei, R. Wobus, J. Ma, B. Cui and S. Moorthi, 2012:

[available at http://www.emc.ncep.noaa.gov/gmb/yzhu/html/imp/201109_imp.html]
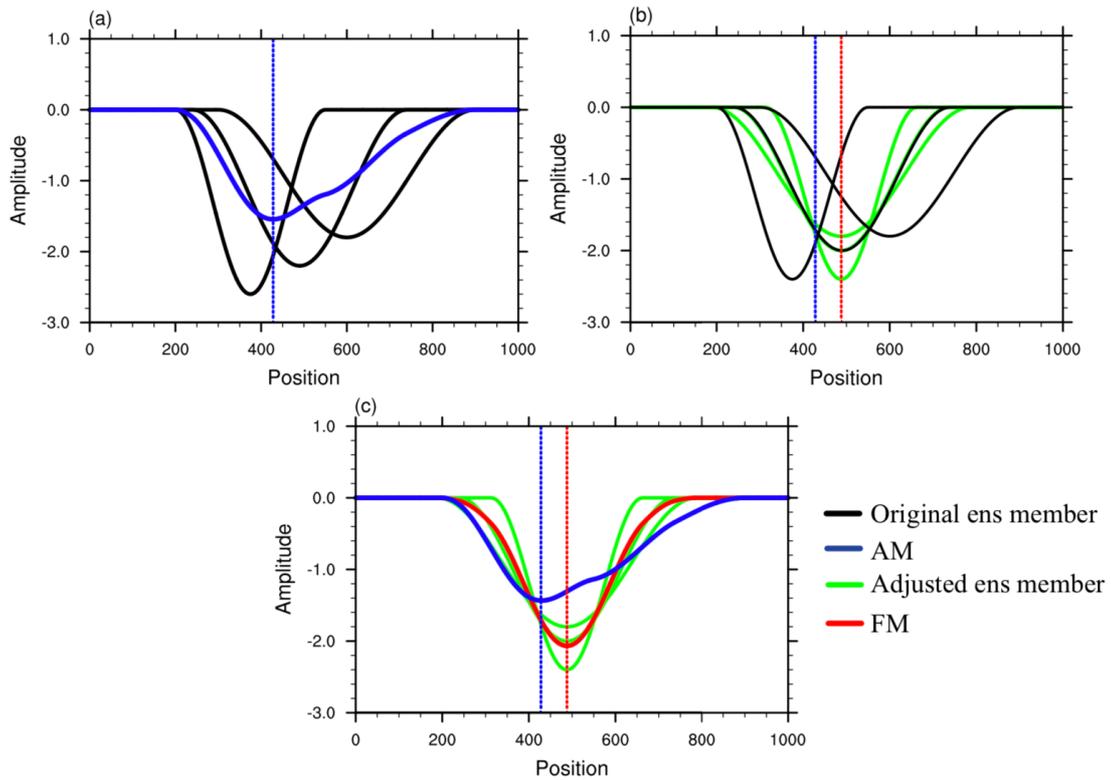
FIG. 1. (a) Schematic of a low-pressure wave (with arbitrary units of amplitude and position) represented in 3 members of an ensemble forecast (solid black lines) and their traditional arithmetic mean (AM, solid blue line). The position of the minimum value of the low-pressure system in AM is marked by a dotted blue vertical line. (b) Schematic of the ensemble members aligned (solid green lines) to the mean of their original position (marked with a dotted red line). (c) The feature-oriented mean (FM, red solid line) is the arithmetic mean of the aligned members. The unit on x and y axises are dimensionless.
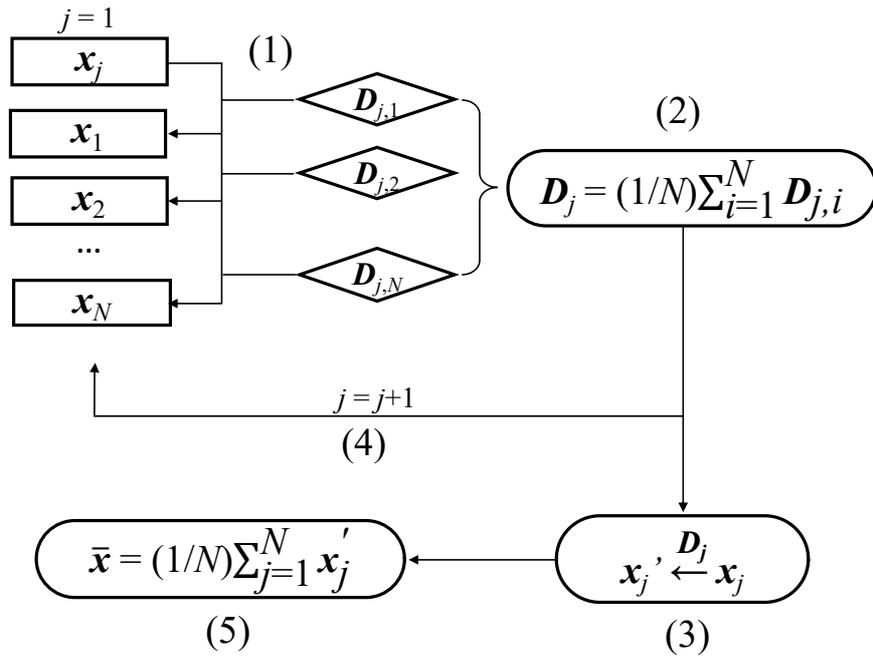
FIG. 2. Flow chart of the FM algorithm.
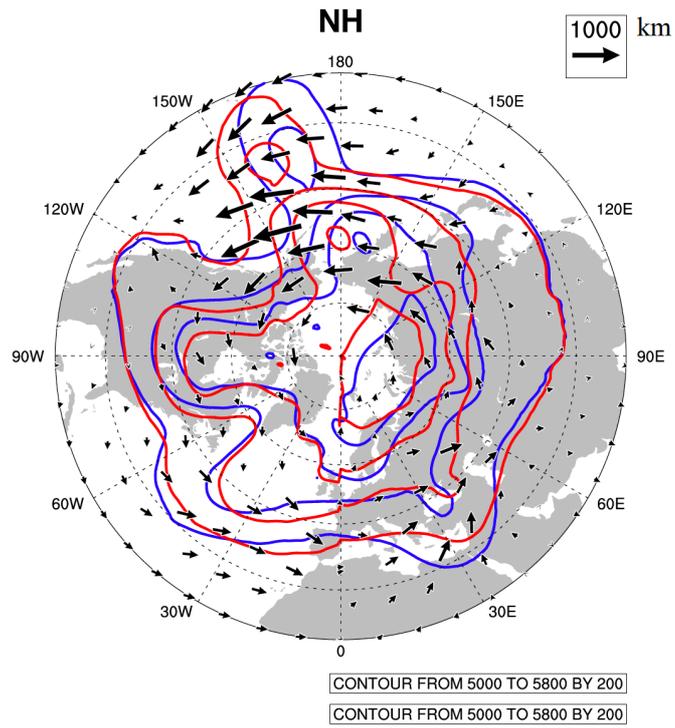
FIG. 3. 500-hPa geopotential height (GH) of a randomly selected 7-day ensemble forecast member (blue contour), its displacement vector (black arrows; unit: km), and the aligned field (red contour) over NH in the case initialized at 0000 UTC 12 Oct 2013.
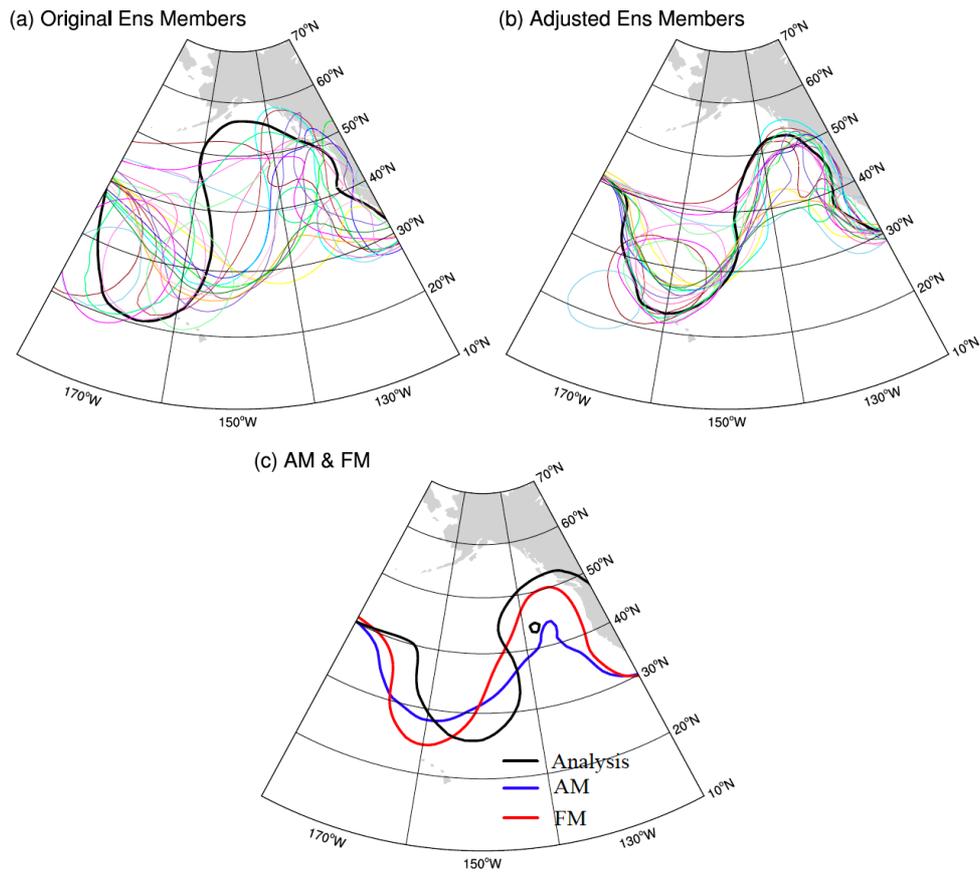
FIG. 4. Spaghetti plot of 500-hPa GH at 5800m of 20 (a) raw and (b) aligned ensemble members in the same case as Fig. 3, and (c) the arithmetic mean of the original (blue, AM) and aligned members (red, FM). For reference, the contour for the verifying analysis is also shown (black).
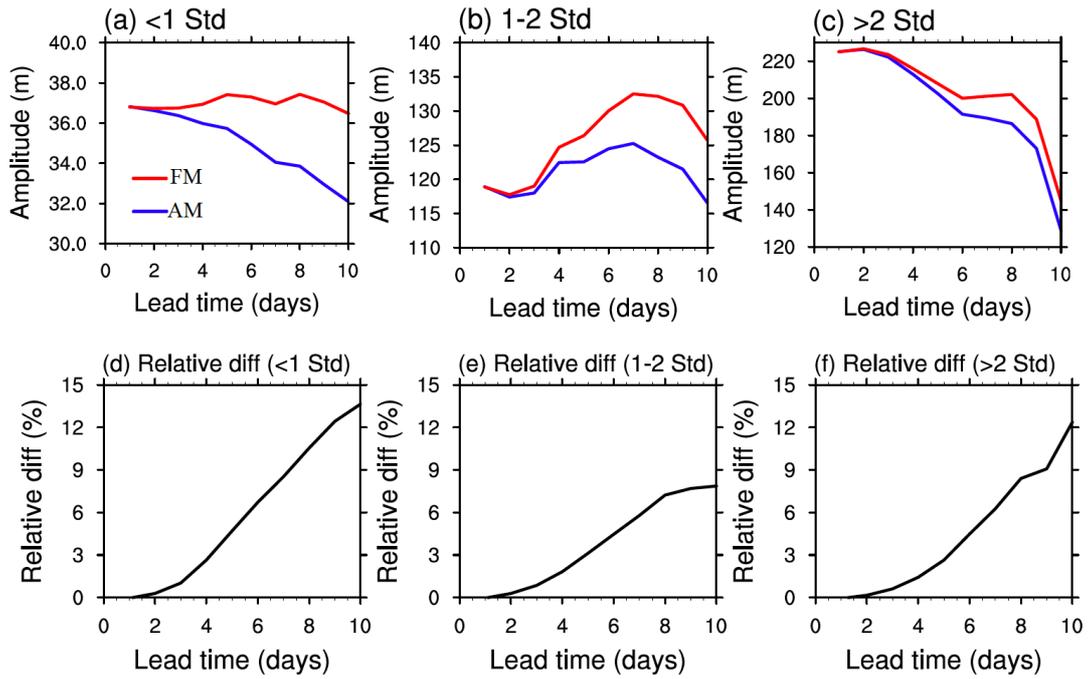
FIG. 5. The temporal and grid mean amplitude of 500-hPa GH between AM (blue) and FM (red) for different categories of events, (a) normal (<1 climatological standard deviation, Std, 80% of cases), (b) medium (1-2 Std, 18% of cases), and (c) extreme (>2 Std, 2% of cases). The categories are divided according to the AM forecast values at each grid point. (d)-(f) are their relative differences, respectively.
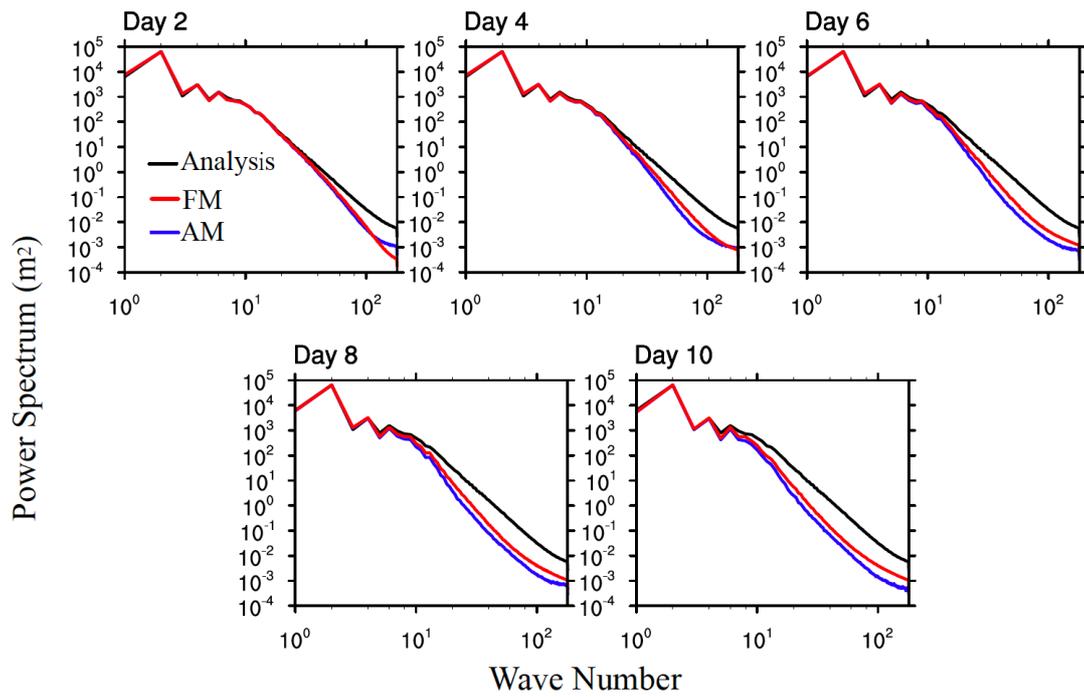
FIG. 6. The spherical harmonic power spectrum of 500-hPa GH for AM (blue), FM (red), and analyses (black) at different lead times.
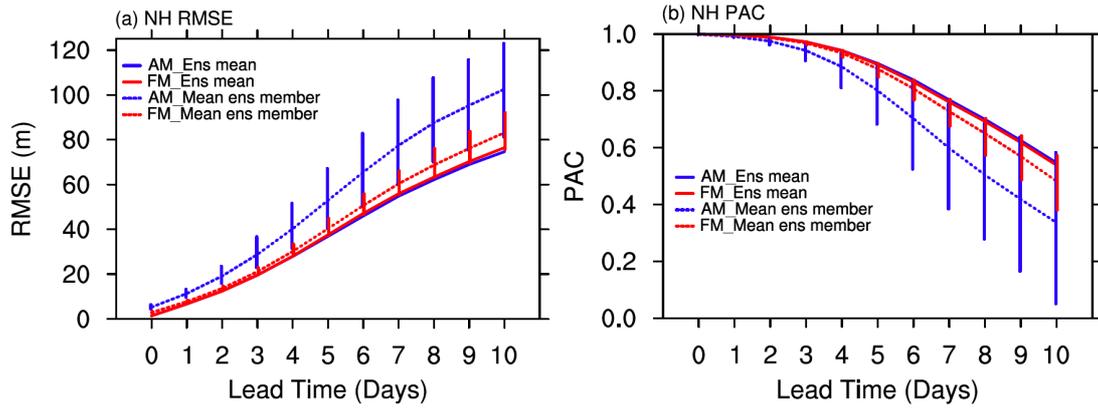
FIG. 7 Temporal mean (a) root-mean-square error (RMSE) and (b) pattern anomaly correlation (PAC) of AM (blue solid line) and FM (red solid line) for 500-hPa GH over the NH extratropic. The mean RMSE and PAC of the original (blue dashed line) and aligned perturbed ensemble forecasts (red dotted line) are also shown, along with vertical bars representing the range of the ensemble values.
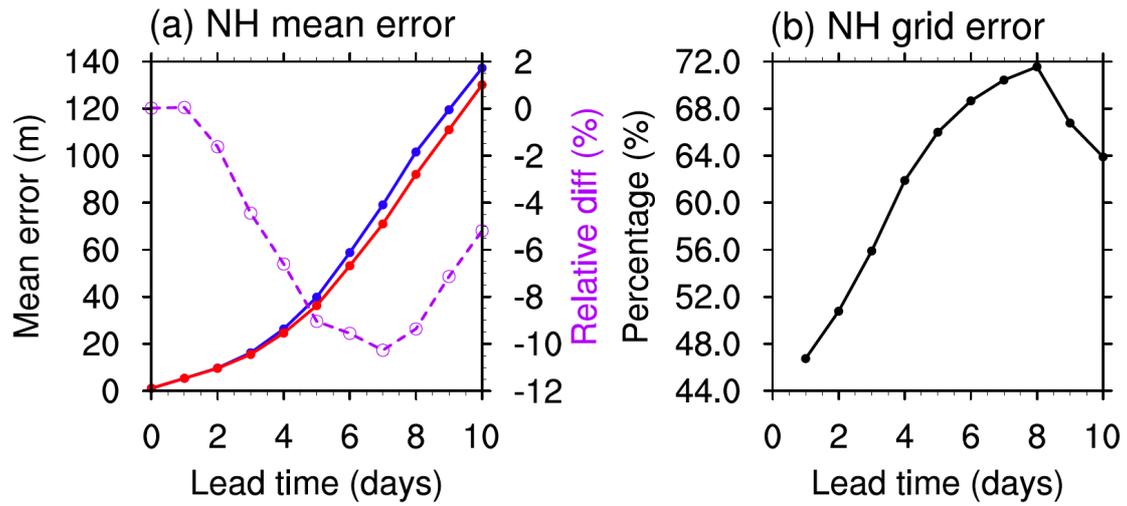
FIG. 8 Temporal and grid mean forecast error of AM (blue) and FM (red) for medium to extreme events of 500-hPa GH over NH. These events are selected for individual grid points with a true value beyond its own 1.5 Std. (b) shows the percentage statistic of grid numbers with a lower error of FM than AM for NH.