A New Measure of Ensemble Central Tendency

JIE FENG

*School of Meteorology, University of Oklahoma, Norman, Oklahoma, and Global Systems Division,
NOAA/Earth System Research Laboratory/Oceanic and Atmospheric Research, Boulder, Colorado*

JING ZHANG AND ZOLTAN TOTH

*Global Systems Division, NOAA/Earth System Research Laboratory/Oceanic and Atmospheric Research,
Boulder, Colorado*

MALAQUIAS PEÑA

Department of Civil and Environmental Engineering, University of Connecticut, Storrs, Connecticut

SAI RAVELA

Earth, Atmosphere and Planetary Sciences, Massachusetts Institute of Technology, Cambridge, Massachusetts

ABSTRACT

Ensemble prediction is a widely used tool in weather forecasting. In particular, the arithmetic mean (AM) of ensemble members is used to filter out unpredictable features from a forecast. AM is a pointwise statistical concept, providing the best sample-based estimate of the expected value of any single variable. The atmosphere, however, is a multivariate system with spatially coherent features characterized with strong correlations. Disregarding such correlations, the AM of an ensemble of forecasts removes not only unpredictable noise but also flattens features whose presence is still predictable, albeit with somewhat uncertain location. As a consequence, AM destroys the structure, and reduces the amplitude and variability associated with partially predictable features. Here we explore the use of an alternative concept of central tendency for the estimation of the expected feature (instead of single values) in atmospheric systems. Features that are coherent across ensemble members are first collocated to their mean position, before the AM of the aligned members is taken. Unlike earlier definitions based on complex variational minimization (field coalescence of Ravela and generalized ensemble mean of Purser), the proposed feature-oriented mean (FM) uses simple and computationally efficient vector operations. Though FM is still not a dynamically realizable state, a preliminary evaluation of ensemble geopotential height forecasts indicates that it retains more variance than AM, without a noticeable drop in skill. Beyond ensemble forecasting, possible future applications include a wide array of climate studies where the collocation of larger-scale features of interest may yield enhanced compositing results.

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 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, Leith 1974) 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 of a single unperturbed forecast (e.g., Toth and Kalnay 1997). Consequently, in the past two decades ensemble mean forecasts became ubiquitous

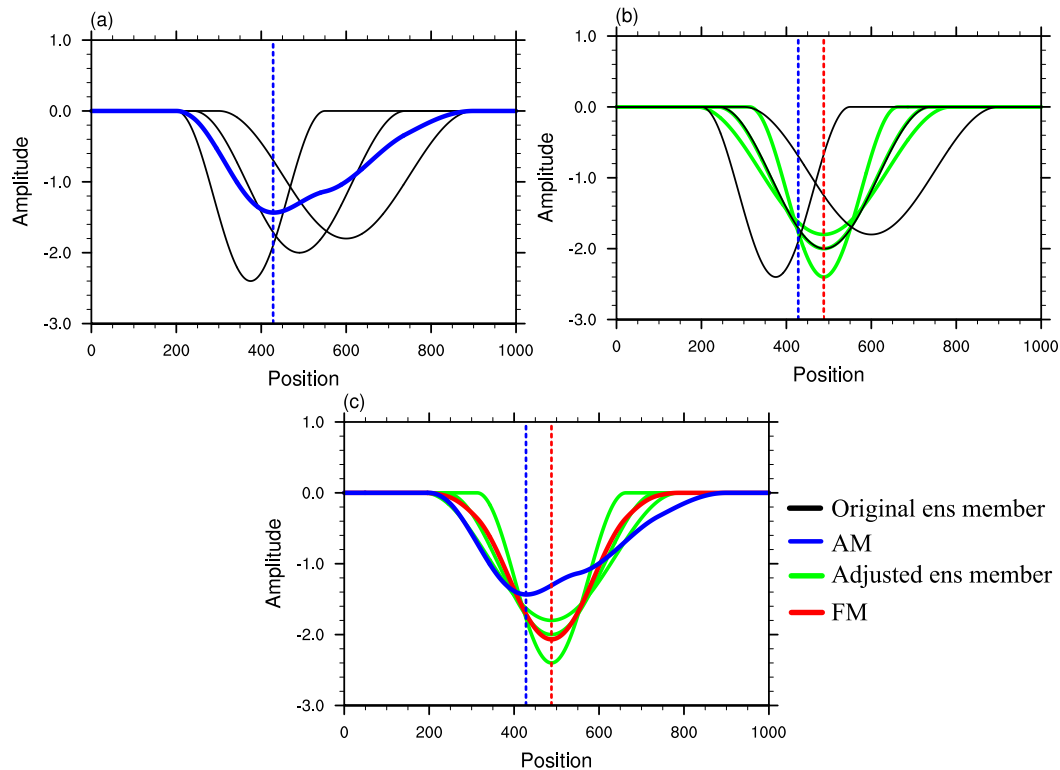*Corresponding author*: Jing Zhang, jing.zhang98567@gmail.com

FIG. 1. (a) Schematic of a low pressure wave (with arbitrary units of amplitude and position) represented in three 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 units on the *x* and *y* axes are dimensionless.

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. As pointed out for example by Molteni et al. (1996), Toth and Kalnay (1997), and Surcel et al. (2014), the removal of unpredictable features in AM results in reduced variability. In particular, the amplitude of features that are partially predictable and thus 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 a sharp low pressure wave present in all ensemble members but at various longitudes as an unrealistically shallow and unrealistically wide low pressure system (see solid blue line in Fig. 1a). Forecast extremes are eliminated and the range of forecast values is reduced, altering the cumulative distribution function (CDF) of AM compared to analysis or individual forecast fields (Ebert 2001). This is a well known problem that 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. These are undesirable characteristics that make AM fields confusing, misleading, and 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:

$$\overline{x_k} = \left(\frac{1}{N}\right)\sum_{i=1}^{N} x_{i,k}, \qquad (1)$$

where $x_{i,k}$ is a single variable at grid point k of the ith member of an N-member ensemble. Clearly, the presence of spatially coherent structures and covariances as in Fig. 1a is not considered in AM.

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), for quasi-normally distributed data, they are statistically nearly identical. 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 misaligned positioning of forecast features.

Recognizing the presence of spatially coherent features in ensemble forecasts, in recent papers, Ravela (field coalescence; Ravela 2012, 2013, 2014) and Purser [generalized ensemble mean (GEM); R. J. Purser (2013, unpublished manuscript)] considered the generalization of central tendency for systems with spatially covarying structures. Note that the new definitions of central tendency given by Ravela (2012) and R. J. Purser (2013, unpublished manuscript) are not explicit but rather operational and use *iterative variational minimization* algorithms. Field coalescence searches for the smallest amplitude two-dimensional (2D) displacement fields that 1) adjust all original ensemble members so their large-scale components are aligned [field alignment (FA) Ravela (2007)], and 2) minimize the difference between individual displaced ensemble fields and their mean. GEM is constructed similarly except for a slightly different variational minimization algorithm.

Both methods require explicit knowledge of the covariance between forecast fields [see Eq. (8) of Ravela (2012), and Eq. 1.5 of R. J. Purser (2013, unpublished manuscript)]. The estimation of such covariances, despite decades of efforts in ensemble-based data assimilation (Hamill et al. 2001; Houtekamer et al. 2005; Wang et al. 2013), remains a challenge. In addition, the calculation of coalesced or generalized ensemble mean fields for operational ensembles with higher model resolution and more ensemble members is computationally rather expensive. Perhaps as a consequence, neither definition has been tested or used in practical weather forecast applications.

Using a similar conceptual approach, here we aim to develop an alternative operational definition for central tendency applicable to fields with spatially coherent structures, and for the first time, apply such a concept in the context of weather forecasting. The new central tendency for ensemble forecasts should filter out the unpredictable, finer-scale features just as AM, while retaining the predictable features with more realistic amplitude/variance. The simplified and computationally efficient method feature-oriented mean (FM) is introduced in section 2. Unlike the methods of Ravela (2012) and R. J. Purser (2013, unpublished manuscript), FM uses a direct, vector-based (i.e., nonvariational) approach that requires no explicit covariance information; instead, it exploits such information implicitly present in the ensemble forecasts. FM is tested using operational ensemble forecasts in section 3. A preliminary evaluation, including a comparison with AM is presented in section 4, while a summary and discussion are offered in section 5.

2. Methodology

As highlighted in the introduction, atmospheric motions manifest in spatiotemporally coherent features of various scales. Yet when comparing states of the atmosphere, a large array of traditional meteorological research and operational applications still use a grid-pointwise approach, disregarding spatiotemporally organized structures. The concept behind FM was pioneered by Ravela (2012, 2013, 2014) and R. J. Purser (2013, unpublished manuscript) and is based on the recognition that the AM used so widely in statistics is a pointwise measure with limited applicability for fields with correlated data. Before taking a pointwise mean of ensemble member forecasts, FM first adjusts all members so their larger-scale features align.

a. Field alignment

The assessment of the spatial displacement and the subsequent adjustment or alignment of meteorological features or fields is a common problem in data assimilation, verification, and other applications. Features can be characterized, for example, by their geographical position, amplitude, or other characteristics (Hoffman et al. 1995; Ravela 2007; Beezley and Mandel 2008). Differences between atmospheric states (e.g., forecast and verifying analysis − 2D forecast error, or unperturbed and perturbed ensemble forecasts − 2D perturbation fields) can be generally partitioned into a positional and a residual (or amplitude) component (e.g., Hoffman et al. 1995; Ravela 2007; Ying 2019; M. Peña, I. Jankov, S. Gregory, S. Ravela, and Z. Toth 2019, unpublished manuscript).

FM uses the FA method (Ravela 2007; Ravela et al. 2007) to establish a 2D mapping between two fields that are similar except some spatial displacement. Let $\mathbf{X}$ and $\mathbf{Y}$ denote 2D fields of system variables (e.g., temperature or geopotential height) with similar but displaced features. For example, $\mathbf{Y}$ is the forecast field, while $\mathbf{X}$ is the targeted field (e.g., the observed or analysis field). FA estimates a smooth 2D displacement vector field $\mathbf{D}$ (see e.g., the black arrows in Fig. 3) that if applied as a translation operation to each grid point of $\mathbf{Y}$ ($\mathbf{Y}$ adjusted to $\mathbf{Y}'$), will minimize the remaining (henceforth called

amplitude) RMS difference between **X** and **Y′**. The displacement vector field **D** and the aligned field **Y′** are derived as the solutions of a variational minimization of the cost function in FA (Ravela 2007). The smoothness of the displacement vector field is controlled by a spectrum truncation algorithm (Ravela 2012). A single tunable wavenumber parameter l determines the scale above which the alignment of features between the two fields is sought, moving the smaller-scale features along with the larger scales, without other adjustments. More details about the FA technique can be found in the above references.

Unlike other techniques used to establish spatial relationships between two fields primarily used in error decomposition studies (e.g., Hoffman et al. 1995; Du et al. 2000; Nehrkorn et al. 2003, 2014), FA uses only one tunable parameter (i.e., the unique smoothing parameter l) and does not rely on posterior (i.e., after alignment) forecast error covariance or any other ancillary information. FA has been used in a wide range of application areas including data assimilation (Ravela et al. 2007), verification (Ravela 2007, 2014; M. Peña, I. Jankov, S. Gregory, S. Ravela, and Z. Toth 2019, unpublished manuscript), nowcasting (Ravela 2012), spatiotemporal error propagation (Feng et al. 2017), among others (Ravela 2015). For easy access by the community, the FA technique was ported into the Developmental Testbed Center (DTC) Code Repository in a study funded by the DTC Visitor Program (PI: S. Ravela).

b. Feature-oriented mean

The key concept behind FM is to align the spatially coherent larger scale and presumably more predictable features in individual ensemble member forecasts to their mean position, while leaving the finer scales that are deemed unpredictable unaligned. As a result, the mean of the aligned, presumably more predictable features in FM retain more the amplitude and the coarser-scale structure of the features in the original member forecasts, while the presumably unpredictable finer-scale, unaligned features are smoothed out in FM as well as in a traditional AM.

Algorithmically, assume that $\mathbf{x}_j$ is a 2D variable field of a randomly selected member of an N-member ensemble $\mathbf{x}_i$ ($i = 1, 2, 3, \ldots, N, i \neq j$). The working definition of FM consists of the following five steps (cf. schematic Fig. 2):

1) Compute the displacement vector $\mathbf{D}_{ji}$ between $\mathbf{x}_j$ and each of the other $N - 1$ members $\mathbf{x}_i$ by using the FA technique with the smoothing parameter l.
2) Calculate the average of displacement vectors $\mathbf{D}_j = (1/N)\sum_{i=1}^{N}\mathbf{D}_{ji}$. The term $\mathbf{D}_j$ represents the displacement
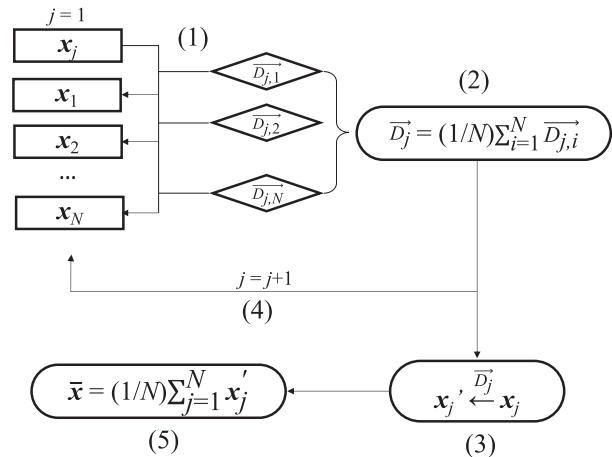

FIG. 2. Flowchart of the FM algorithm.

of features in $\mathbf{x}_j$ compared to the mean of the position of features in all ensemble members.
3) Adjust member $\mathbf{x}_j$ by transposing its 2D field in space by the displacement vector $\mathbf{D}_j$: $\mathbf{x}_j' \leftarrow \mathbf{x}_j + \mathbf{D}_j$. This will align the position of features in member $\mathbf{x}_j$ to the mean of their position in the entire ensemble.
4) Repeat steps 1–3 for each member of the ensemble. The aligned members (illustrated by the green lines in Fig. 1b) will differ only in the amplitude of their features, apart from inconsistencies (i.e., unpredictable noise) on the smaller scales that are unaffected by the FA procedure.
5) FM is defined as the arithmetic mean of all aligned ensemble members: $\overline{\mathbf{x}} = (1/N)\sum_{j=1}^{N}\mathbf{x}_j'$ (see red solid line in Fig. 1c).

Note that the adjusted individual ensemble members (i.e., $\mathbf{x}_j'$) are intended only for the derivation of FM, not for other applications.

As pointed out earlier, FM is conceptually very similar to Ravela's (2012) field coalescence and R. J. Purser's (2013, unpublished manuscript) GEM methods. Both FM and field coalescence calculate the mean amplitude field by aligning features in each ensemble field to the mean of their position in the individual fields, using the FA technique of Ravela (2007) as an alignment tool. Field coalescence, however, solves a more complex variational minimization problem to estimate the mean amplitude field and displacement fields for each member at once. In contrast, FM solves a set of much simpler FA minimization problems followed by a simplified vector calculation to derive an estimate for mean position, after which the original members are transposed with their corresponding displacement vector fields. Moreover, unlike both the field coalescence and GEM algorithms, FM does not require covariance information about the state being estimated.

3. Experimental setup

The FM algorithm described in section 2b will be evaluated and compared to AM using forecasts from the National Centers for Environmental Prediction (NCEP) Global Ensemble Forecasting System (GEFS; Toth and Kalnay 1993, 1997; Zhu et al. 2012; Zhou et al. 2016, 2017). The 20-member 500-hPa geopotential height (GH) 0000 UTC GEFS forecasts from a 25-day sample period (1 October–25 October 2013) will be compared to corresponding verifying analysis fields from the Global Forecast System (GFS) on a common 1° horizontal resolution grid. Ensemble forecasts offer an ideal testing ground for the use of FM due to the coherence of larger-scale and more predictable features across ensemble members that FM can readily detect.

Ideally, displacement vectors could be defined in 4D (3D in space and 1D in time). Note that at the time of writing, only a 2D version of FA is available. As the amount and complexity of horizontal variability much exceed that in the vertical, the approximation of the 2D application may only have a small effect. As for the time dimension, as long as ensemble perturbations are coherent in 4D, 2D horizontal plane alignments will necessarily lead to the alignment of features in time as well.

The only free parameter in FM is the smoothing parameter l of FA. Ideally, one would want to choose this parameter so the displacement vectors reflect differences in the position of the partially predictable features. This is feasible via a lead time dependent specification of l, exploiting the strong dependence of predictability on scales (Boer 2003). For simplicity, in the initial tests presented in section 4, for all lead times, l is set to be 128 (~300-km wavelength), covering part of the mesoscale and larger-scale systems.

4. Results

After the visual demonstration in a real case scenario, FM will be quantitatively evaluated by assessing (i) how much more variance is retained by FM versus AM [by comparing the wavenumber spectrum of AM and FM with that of the analysis fields, e.g., Adams and Swarztrauber (1997)]; and (ii) how much, if any, the performance of FM is degraded compared to AM [pattern anomaly correlation (PAC) and RMSE; e.g., Buizza et al. (2005)].

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°N) forecast initialized at 0000 UTC 12 October 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). The short displacement vectors and correspondingly small displacement of contour lines over the Atlantic indicate that in that region, features in the selected ensemble member (member 10) align well with the mean position of features in the rest of the ensemble members. On the contrary, the selected ensemble member appears to be an outlier over the Pacific, as manifested by the long displacement vectors and correspondingly large displacement of contour lines around 160°W. This is confirmed by Fig. 4a where the heavy magenta 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 position of the features 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 forecasting or climatological compositing 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. The results (Fig. 5a) show that the mean amplitude of FM and AM maintains lower than that of analysis and the differences expand with longer lead times. It is because more scales of features in forecasts become less predictable and are thus smoothed as the lead time increases. However, the amplitude of FM remains higher than that of AM for all lead times (Fig. 5a). The relative improvement of FM using analysis as a reference monotonically increases to about 10% at day 10 (Fig. 5b).

The spectral distribution of analysis and various lead time AM (blue) and FM (red) forecasts averaged over 25 cases is displayed in Fig. 6. Probably due to the initial approximately linear evolution of perturbations (Gilmour et al. 2001), at day 2, only minor differences and only on the finest scales are seen between the two types of mean forecasts. As with increasing lead time nonlinearities emerge on larger scales, the variance in AM falls below that of natural variability in the analysis (black), and at progressively larger scales. FM, meanwhile, retains more natural variability on those scales, due to the alignment of larger-scale features before the mean of the member forecasts is taken. Statistically
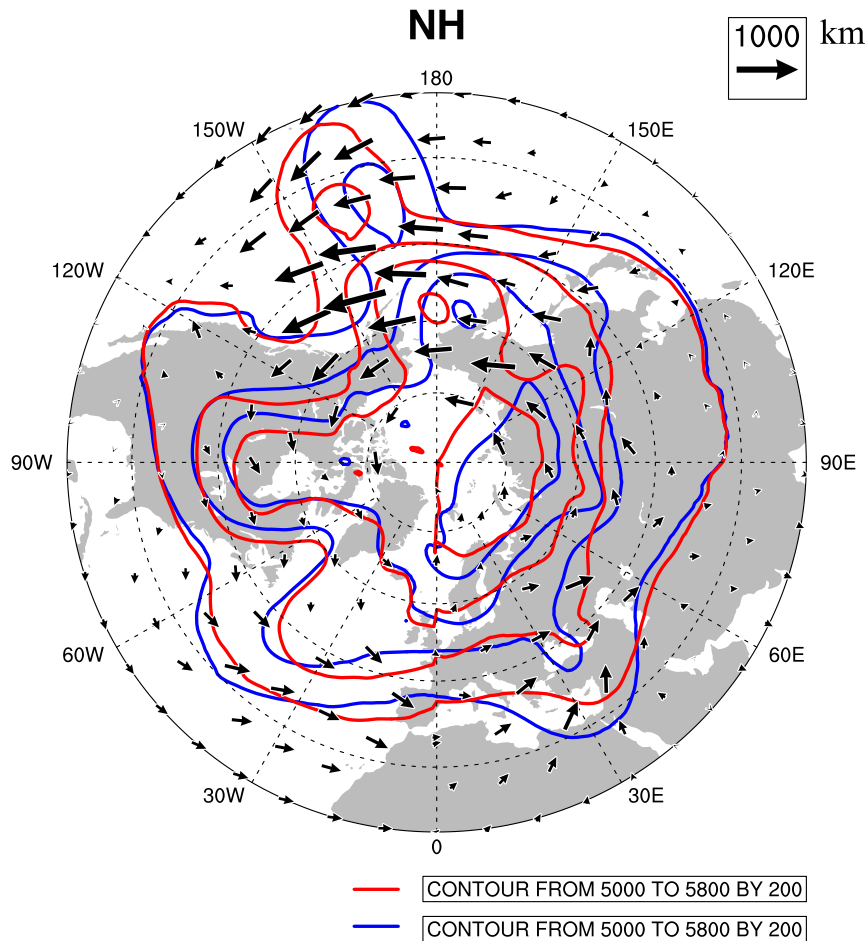
NH

FIG. 3. The 7-day Northern Hemisphere extratropical (NH; 20°–80°N) 500-hPa geopotential height (GH) forecast from a randomly selected ensemble member (blue contours, initialized at 0000 UTC 12 Oct 2013), its displacement vector field (black arrows;km), and the field aligned with the mean of the position of features in all ensemble members (red contours).

significant differences (according to a standard t test at the 0.05 significance level) first appear on fine scales, extending with increasing lead time to larger, up to wavenumber-7 scales. Nonetheless, the difference between FM and AM variance is relatively small compared to their distinction from the analysis. The variance in FM is expected to be below that of analysis fields due to the removal of finer-scale features that remain fully out of phase even after the alignment of the predictable scales.

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 (e.g., Leith 1974; Li et al. 2018; Feng et al. 2019). This is due to the (sample size dependent) reduction of natural variance (i.e., noise) in the sample. Any deviation from the AM formula in Eq. (1) can only

increase RMSE in the estimate of the expected value. How much FM may increase the error found in AM while retaining larger amplitudes (Fig. 5) and more variance (Fig. 6)? Figure 7 illustrates the mean of 500-hPa NH geopotential height RMSE and ensemble spread (Fig. 7a), and PAC (Fig. 7b) for AM and FM. As the proximity of the blue (AM) and red (FM) lines indicate, forecast performance is only slightly affected by the retention of more variance in FM. Results for the SH (not shown) are similar.

Note that the alignment of members is tantamount to the elimination of positional perturbations. As seen from Fig. 7, this leads to an increase in the skill of the aligned members (dashed red) compared to the original members (dashed blue), closing much of the skill gap present between the original perturbed and AM (solid blue) forecasts. At the same time, FM deceases more than 50% the spread of members around their respective
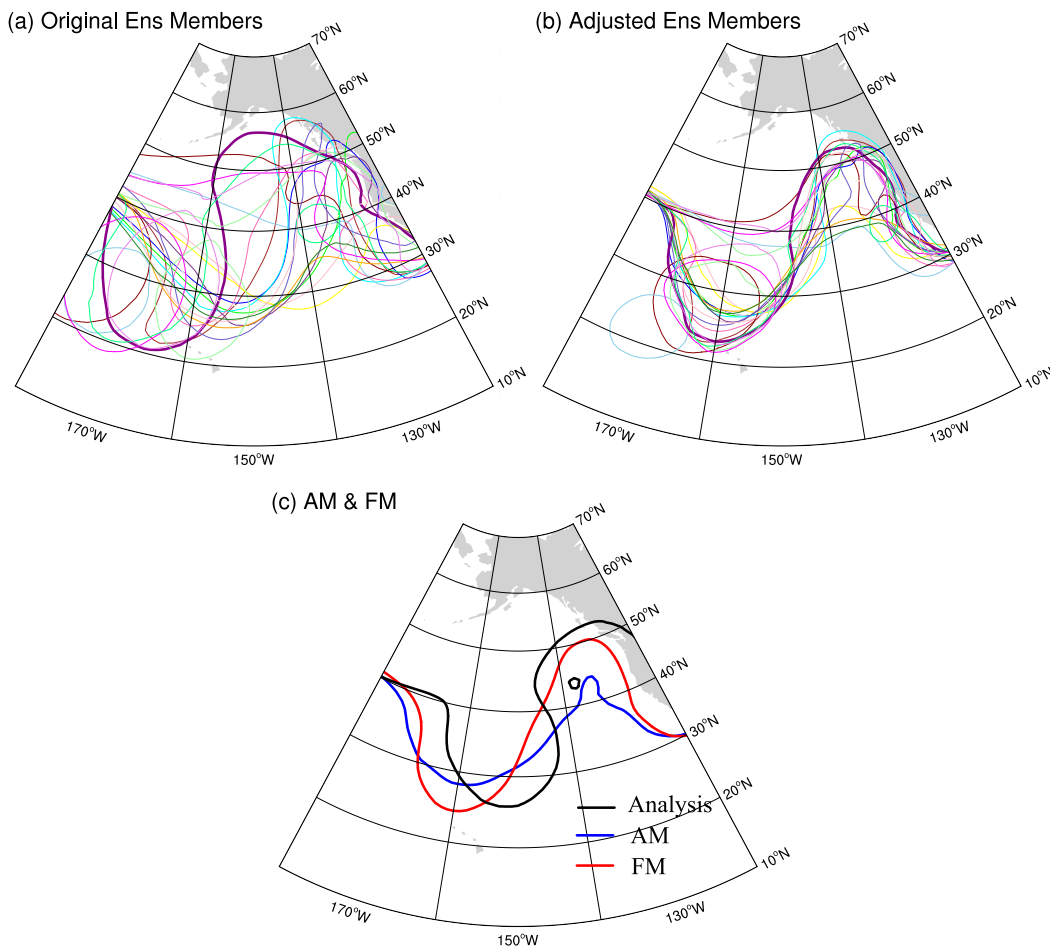
FIG. 4. Spaghetti plot for the 5800-m contour of the 500-hPa GH forecast displayed in Fig. 2 for (a) 20 raw and (b) aligned ensemble members, and (c) the arithmetic mean of the original (blue; AM) and aligned members (red; called FM). For reference, the contour for the verifying analysis is also shown (black).

means (FM for aligned members, solid red, or AM for original members, solid blue). The much reduced spread of the aligned members around FM reflects the uncertainty only in the amplitude or structure (but not the position) of forecast features.

Interestingly, in cases with strong observed anomalies (i.e., where the verifying analysis has an anomaly larger than 1.5 climatological standard deviation), FM (red line in Fig. 8) has a lower RMSE than AM (blue line). The error reduction peaks at 10% around day 7, with FM error lower than AM error at 70% of all NH grid points. The results for the SH (not shown) are similar.

With $l = 128$, FM is expected to perform best when synoptic scales (~1000 km) become partially predictable. This is consistent with results presented in Figs. 5–8, showing the strongest impact beyond day 4. Shorter lead times are characterized with quasi-linear evolution of ensemble perturbations so FM and AM have small differences. At extended ranges (e.g., 10 days and beyond),

nonlinear saturation has a strong influence on the evolution of perturbations even at the largest scales. As forecast skill diminishes, features across ensemble members become uncorrelated, preventing the variational FA algorithm in step (1) of the FM algorithm to converge to a displacement vector field solution. Consequently, FM may no longer be applicable at these extended ranges.

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 that are out of phase across members). Not surprisingly, AM gained widespread use in weather forecasting. On the other hand, when applied in a traditional, pointwise sense, AM reduces the amplitude, and distorts the structure of partly predictable features that are present in the perturbed ensemble forecasts at
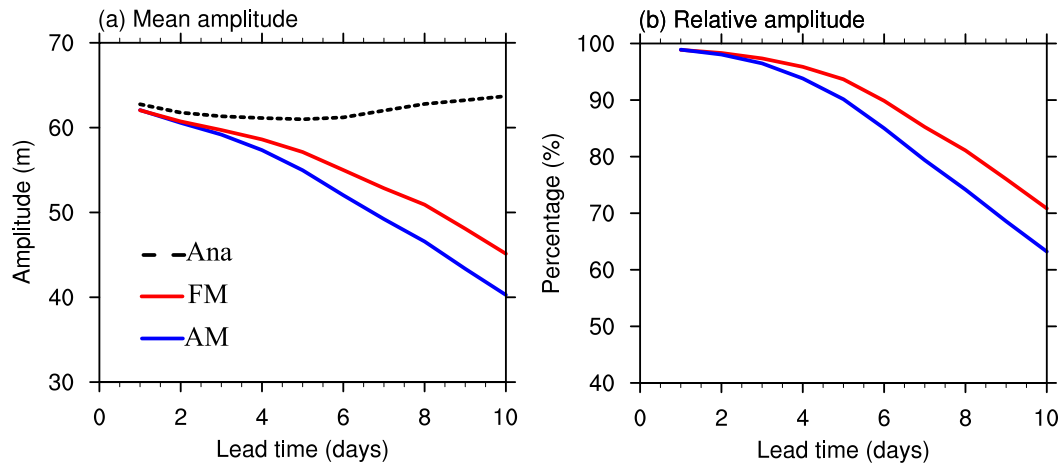
FIG. 5. (a) Global and sample (1–25 Oct 2013) mean amplitude of 500-hPa GH analysis (black dashed), AM (blue), and FM (red) anomalies against the climatological mean as a function of lead time. (b) Amplitude percentage of AM and FM relative to analysis as a function of lead time.

somewhat different locations and with somewhat varying structures. Recognizing the spatiotemporal coherence of partly predictable features across ensemble members, we propose to spatially collocate such features before their mean is taken. In the new ensemble

central tendency that we call feature-oriented mean (FM), all larger-scale 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. Instead of
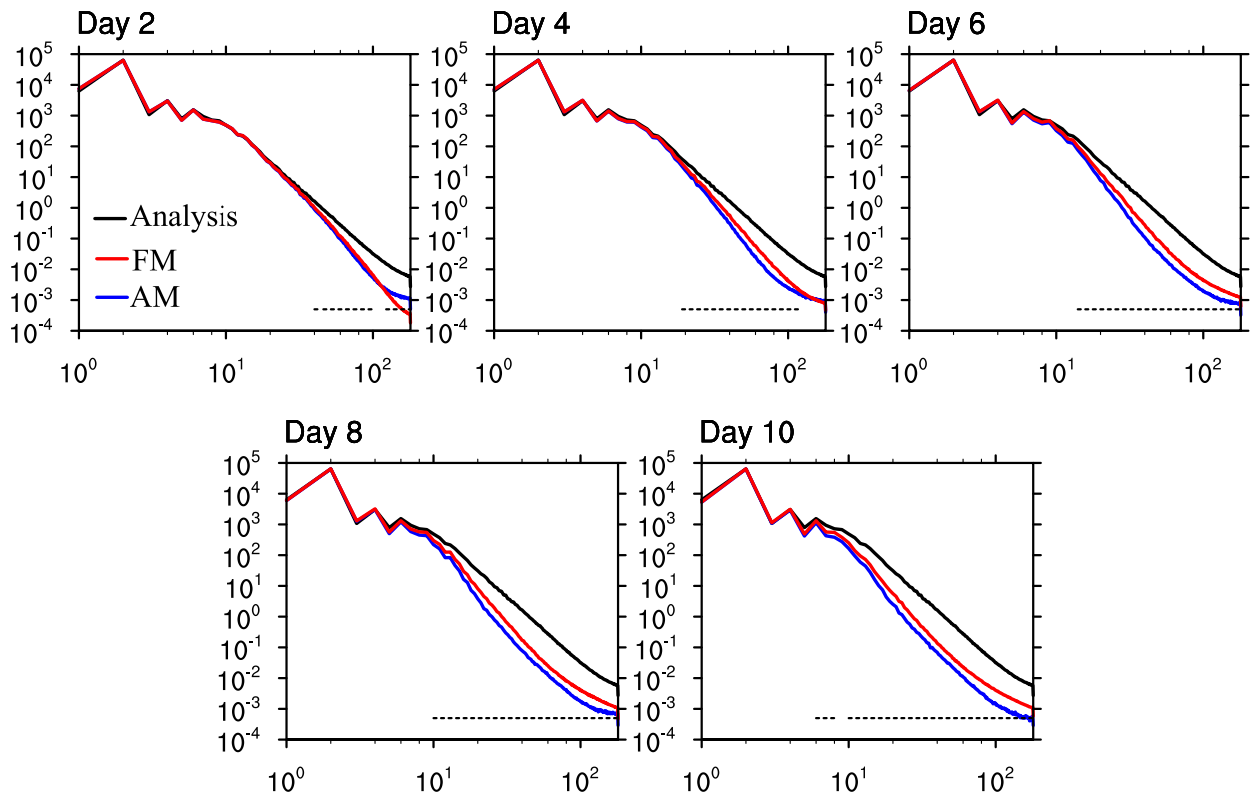
FIG. 6. The spherical harmonic power spectrum of 500-hPa GH over the globe for AM (blue) and FM (red) forecasts, and analyses (black solid lines) at different lead times. The black dashed lines indicate the scales at which the spectral variance of FM and AM are statistically different at the 5% significance level.
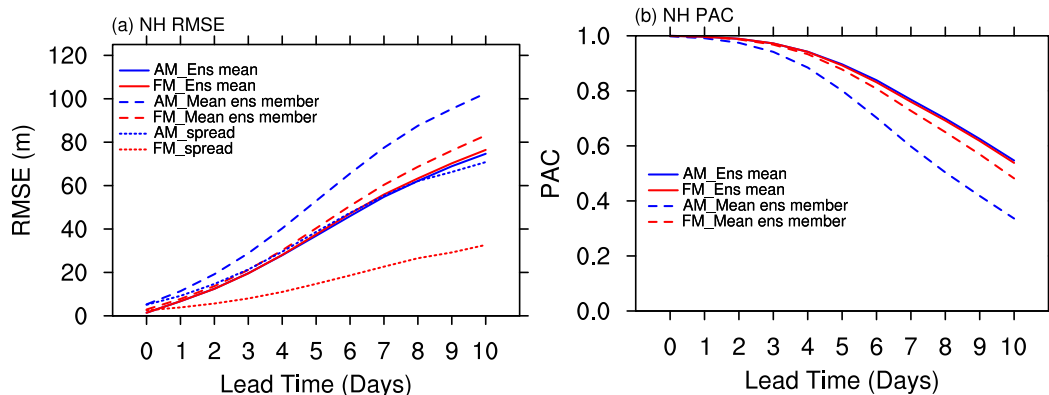
FIG. 7. Temporal mean (1–25 Oct 2013) (a) root-mean-square error (RMSE; solid lines) and spread (dotted lines) and (b) pattern anomaly correlation (PAC; solid lines) for AM (blue lines) and FM (red lines) for 500-hPa GH over the NH extratropics. The mean of the RMSE and PAC of the original (blue dashed lines) and aligned perturbed ensemble forecasts (red dashed lines) are also shown.

averaging a collection of variables to estimate the *expected value* of assumedly uncorrelated spatial grids like in AM, FM estimates the *expected feature* of system variables, composed of features with covarying elements.

FM is based on the same concepts as central tendencies proposed by Ravela (2012; field coalescence) and R. J. Purser (2013, unpublished manuscript; GEM). However, these methods have not found their way into meteorological applications, probably due to their complexity and high computational demand. In contrast, FM is defined with a simplified and computationally efficient method, via a direct vector-based calculation. Particularly, FM does not require the amplitude covariance information that, as a prior estimation, is critical in the coalescence and GEM methods.

On the computational side, the bulk of the FM algorithm (steps 1–4 in section 2b) 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 high-quality first guess for FA calculations at the next lead time.

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, resulting in features with more realistic amplitude. Meanwhile, forecast performance is not compromised as FM RMSE and PAC is hardly changed compared to
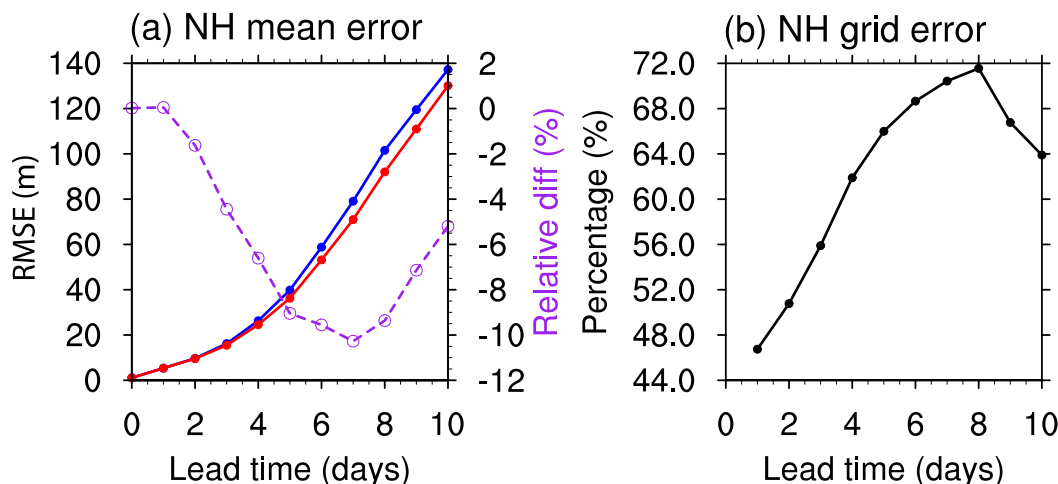

FIG. 8. (a) AM and FM RMSE (solid lines) as in Fig. 7a, but restricted for grid points where the anomaly of the verifying analysis is greater than 1.5 times the climatological standard deviation. The dashed curve represents the difference between the two error curves as a percentage of AM error. (b) Percentage of grid points where the absolute error in FM is lower than in AM.

AM. Interestingly, FM outperforms AM under extreme observed conditions, particularly for medium-range (i.e., 3–8-day lead time synoptic scale) forecasts.

For simplicity, the smoothing coefficient in the reported FA experiments was fixed over all lead times. Ideally, as a function of lead time, one would increase the level of smoothing (i.e., numerically lower the smoothing parameter l), making the displacement vector field to initially reflect most, while later only coarser scales, in accordance with the upscale propagation of forecast errors with increasing lead time. At long lead times, all predictability is lost and ensemble members constitute a random draw from climatology (i.e., no coherency in structures across ensemble members). Smoothing can then approach its maximum level (i.e., only a single vector, or no spatial adjustment at all over the entire domain), at which stage FM asymptotes to AM. As is the case with any forecast, a reduction or elimination of model systematic errors may further improve the performance of FM.

Though FM was demonstrated with 500-hPa height data, it can be applied to any variable of interest, including noncontinuous variables such as precipitation.[1] 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 a multimodel superensemble (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 collocate and then average any feature of interest.

REFERENCES

Adams, J. C., and P. N. Swarztrauber, 1997: Spherepack 2.0: A model development facility. NCAR Tech. Note NCAR/TN-436-STR, 62 pp., https://doi.org/10.5065/D6Z899CF.

Beezley, J. D., and J. Mandel, 2008: Morphing ensemble Kalman filters. *Tellus*, **60A**, 131–140, https://doi.org/10.1111/j.1600-0870.2007.00275.x.

Boer, G. J., 2003: Predictability as a function of scale. *Atmos.–Ocean*, **41**, 203–215, https://doi.org/10.3137/ao.410302.

Buizza, R., P. L. Houtekamer, Z. Toth, G. Pellerin, M. Wei, and Y. Zhu, 2005: A comparison of the ECMWF, MSC, and NCEP global ensemble prediction systems. *Mon. Wea. Rev.*, **133**, 1076–1097, https://doi.org/10.1175/MWR2905.1.

Delle Monache, L., X. Deng, Y. Zhou, and R. Stull, 2006: Ozone ensemble forecasts: 1. A new ensemble design. *J. Geophys. Res.*, **111**, D05307, https://doi.org/10.1029/2005JD006310.

Du, J., S. L. Mullen, and F. Sanders, 2000: Removal of distortion error from an ensemble forecast. *Mon. Wea. Rev.*, **128**, 3347–3351, https://doi.org/10.1175/1520-0493(2000)128<3347:RODEFA>2.0.CO;2.

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, https://doi.org/10.1175/1520-0493(2001)129<2461:AOAPMS>2.0.CO;2.

Feng, J., Z. Toth, and M. Peña, 2017: Spatially extended estimates of analysis and short-range forecast error variances. *Tellus*, **69A**, 1325301, https://doi.org/10.1080/16000870.2017.1325301.

——, J. Li, J. Zhang, D. Liu, and R. Ding, 2019: The relationship between deterministic and ensemble mean forecast errors revealed by global and local attractor radii. *Adv. Atmos. Sci.*, **36**, 271–278, https://doi.org/10.1007/s00376-018-8123-5.

Galmarini, S., and Coauthors, 2004: Ensemble dispersion forecasting—Part I: Concept, approach and indicators. *Atmos. Environ.*, **38**, 4607–4617, https://doi.org/10.1016/j.atmosenv.2004.05.030.

Gilmour, I., L. A. Smith, and R. Buizza, 2001: Linear regime duration: Is 24 hours a long time in synoptic weather forecasting? *J. Atmos. Sci.*, **58**, 3525–3539, https://doi.org/10.1175/1520-0469(2001)058<3525:LRDIHA>2.0.CO;2.

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, https://doi.org/10.1175/1520-0493(2001)129<2776:DDFOBE>2.0.CO;2.

Hoffman, R. N., Z. Liu, J. Louis, and C. Grassotti, 1995: Distortion representation of forecast errors. *Mon. Wea. Rev.*, **123**, 2758–2770, https://doi.org/10.1175/1520-0493(1995)123<2758:DROFE>2.0.CO;2.

---

[1] A continuous variable related to precipitation can be used. For example, the pseudo precipitation (Yuan et al. 2019) combines information from the spatially and temporally continuous precipitable water and the discontinuous precipitation variables.

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, https://doi.org/10.1175/MWR-2864.1.

Knutti, R., R. Furrer, C. Tebaldi, and J. Cermak, 2010: Challenges in combining projections from multiple climate models. *J. Climate*, **23**, 2739–2758, https://doi.org/10.1175/2009JCLI3361.1.

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, https://doi.org/10.1002/2015RG000513.

Leith, C. E., 1974: Theoretical skill of Monte Carlo forecasts. *Mon. Wea. Rev.*, **102**, 409–418, https://doi.org/10.1175/1520-0493(1974)102<0409:TSOMCF>2.0.CO;2.

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, https://doi.org/10.1007/s00382-017-4017-y.

Lorenz, E. N., 1963: Deterministic nonperiodic flow. *J. Atmos. Sci.*, **20**, 130–141, https://doi.org/10.1175/1520-0469(1963)020<0130:DNF>2.0.CO;2.

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, https://doi.org/10.1002/qj.49712252905.

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, https://doi.org/10.1256/qj.02.18.

——, 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, https://doi.org/10.1175/MWR-D-13-00164.1.

Ravela, S., 2007: Two extensions of data assimilation by field alignment. *Computational Science— ICCS 2007(ICCS 2007)*, Y. Shi et al., Eds., Vol. 4487, *Lecture Notes in Computer Science*, Springer, 1147–1154, https://doi.org/10.1007/978-3-540-72584-8_150.

——, 2012: Quantifying uncertainty for coherent structures. *Procedia Comput. Sci.*, **9**, 1187–1196, https://doi.org/10.1016/j.procs.2012.04.128.

——, 2013: Generating a forecast by field coalescence. U.S. patent application US 13/904, 058, https://patents.google.com/patent/US9046627B2/en?inventor=ravela&oq=ravela.

——, 2014: Spatial inference for coherent geophysical fluids by appearance and geometry. *Proc. IEEE Winter Conf. on Applications of Computer Vision*, Steamboat Springs, CO, IEEE, 925–932, https://doi.org/10.1109/WACV.2014.6836005.

——, 2015: Statistical inference for coherent fluids. *Dynamic Data-Driven Environmental Systems Science (DyESS 2014)*, S. Ravela and A. Sandu, Eds., Vol. 8964, *Lecture Notes in Computer Science*, Springer, 121–133, https://doi.org/10.1007/978-3-319-25138-7_12.

——, K. Emanuel, and D. McLaughlin, 2007: Data assimilation by field alignment. *Physica D*, **230**, 127–145, https://doi.org/10.1016/j.physd.2006.09.035.

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, https://doi.org/10.1175/MWR-D-13-00134.1.

Toth, Z., and E. Kalnay, 1993: Ensemble forecasting at NMC: The generation of perturbations. *Bull. Amer. Meteor. Soc.*, **74**, 2317–2330, https://doi.org/10.1175/1520-0477(1993)074<2317:EFANTG>2.0.CO;2.

——, and ——, 1997: Ensemble forecasting at NCEP and the breeding method. *Mon. Wea. Rev.*, **125**, 3297–3319, https://doi.org/10.1175/1520-0493(1997)125<3297:EFANAT>2.0.CO;2.

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, https://doi.org/10.1175/MWR-D-12-00141.1.

Ying, Y., 2019: A multiscale Alignment method for ensemble filtering with displacement errors. *Mon. Wea. Rev.*, **147**, 4553–4565, https://doi.org/10.1175/MWR-D-19-0170.1.

Yuan, H., Z. Toth, M. Peña, and E. Kalnay, 2018: Overview of weather and climate systems. *Handbook of Hydrometeorological Ensemble Forecasting*, Q. Duan et al., Eds., Springer, 35–65.

——, P. Schultz, E. Tollerud, D. Hou, Y. Zhu, M. Pena, M. Charles, and Z. Toth, 2019: Pseudo-precipitation: A continuous precipitation variable. NOAA Tech. Memo. OAR GSD-62, 15 pp., https://doi.org/10.25923/3h37-gp49.

Zhou, B., and J. Du, 2010: Fog prediction from a multimodel mesoscale ensemble prediction system. *Wea. Forecasting*, **25**, 303–322, https://doi.org/10.1175/2009WAF2222289.1.

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, https://doi.org/10.1175/WAF-D-16-0109.1.

——, ——, ——, 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, https://doi.org/10.1175/WAF-D-17-0023.1.

Zhu, Y., D. Hou, M. Wei, R. Wobus, J. Ma, B. Cui, and S. Moorthi, 2012: February 2012 upgrade of the NCEP Global Ensemble Forecast System (NAEFS). NOAA, 29 pp., https://www.emc.ncep.noaa.gov/gmb/yzhu/imp/i201109/GEFS_Science_20120208.pdf.