Supplement of Nonlin. Processes Geophys., 31, 287–302, 2024
https://doi.org/10.5194/npg-31-287-2024-supplement

Nonlinear Processes in Geophysics
Open Access

Supplement of

Improving ensemble data assimilation through Probit-space Ensemble Size Expansion for Gaussian Copulas (PESE-GC)

Man-Yau Chan

Correspondence to: Man-Yau Chan (chan.1063@osu.edu)

S1. Ensemble modulation results in non-Gaussian expanded ensemble statistics

Ensemble modulation constructs an expanded ensemble by combining a localization matrix with forecast ensemble perturbations (Bishop and Hodyss, 2009, 2011; Bishop et al., 2017). This is done by taking Schur products (i.e., element-wise products) between 1) forecast ensemble perturbations and 2) the columns of a square-root of the localization matrix. In this appendix, it will be shown that if the forecast ensemble has Gaussian statistics, the expanded ensemble (henceforth, the modulated ensemble) is likely to possess non-Gaussian statistics.

Suppose the localization matrix $\mathbf{\Phi}$ is an $N_x \times N_x$ positive semi-definite symmetric matrix with rank N_L, $\boldsymbol{L}$ is an $N_x \times N_L$ matrix square root of the localization matrix (i.e., $\mathbf{\Phi} = \boldsymbol{L}\boldsymbol{L}^\top$), $N_x \gg N_E$ and $N_x \geq N_L$. Supposing ℓ_m is the m-th column of $\boldsymbol{L}$ and $\boldsymbol{x'_n}$ is the n-th forecast ensemble perturbation, a modulated ensemble perturbation $\boldsymbol{v'_k}$ $(k \equiv mn)$ can be created from ℓ_m and $\boldsymbol{x'_n}$ by (Bishop et al., 2017)

$$\boldsymbol{v'_k} \equiv \sqrt{\frac{N_E N_L}{N_E - 1}} \boldsymbol{\ell_m} \circ \boldsymbol{x'_n} \;\; \forall k = 1, 2, \ldots, N_E N_L \tag{S1}$$

where $\circ$ represents the element-wise product. For simplicity, suppose $N_E \gg 1$. Then,

$$\boldsymbol{v'_k} \approx \sqrt{N_L} \boldsymbol{\ell_m} \circ \boldsymbol{x'_n} \;\; \forall k = 1, 2, \ldots, N_E N_L \tag{S2}$$

To show that $\boldsymbol{v'_k}$ has non-Gaussian statistics, consider the moments of some g-th element in the modulated ensemble perturbation vector. Supposing $v'_{k,g}$ is the g-th element of $\boldsymbol{v'_k}$, the p-th central moment of $v'_{k,g}$ can be written as

$$\left\langle \left(v'_g\right)^p \right\rangle \equiv \frac{1}{N_E N_L} \sum_{m=1}^{N_L} \sum_{n=1}^{N_E} v'_{k,g} \tag{S3}$$

For simplicity, assume N_E is sufficiently large such that the above expression is an approximately unbiased estimator for $\left\langle \left(v'_g\right)^p \right\rangle$. Substituting Eq. (S2) into Eq. (S3) gives

$$\begin{aligned} \left\langle \left(v'_g\right)^p \right\rangle &\approx \frac{1}{N_E N_L} \sum_{m=1}^{N_L} \sum_{n=1}^{N_E} \left(\ell_{m,g} x'_{n,g} \sqrt{N_L} \right)^p \\ &= N_E^{-1} N_L^{p/2-1} \sum_{m=1}^{N_L} \sum_{n=1}^{N_E} \left(\ell_{m,g} x'_{n,g} \right)^p \end{aligned} \tag{S4}$$

where $\ell_{m,g}$ is the g-th element of $\boldsymbol{\ell_m}$ and $x'_{n,g}$ is the g-th element of $\boldsymbol{x'_n}$. Applying some algebraic manipulation yields

$$\begin{aligned} \left\langle \left(v'_g\right)^p \right\rangle &\approx N_L^{p/2-1} \left[\sum_{m=1}^{N_L} \left(\ell_{m,g}\right)^p \right] \left[\frac{1}{N_E} \sum_{n=1}^{N_E} \left(x'_{n,g}\right)^p \right] \\ &= N_L^{p/2-1} \left[\sum_{m=1}^{N_L} \left(\ell_{m,g}\right)^p \right] \left\langle \left(x'_g\right)^p \right\rangle \end{aligned} \tag{S5}$$

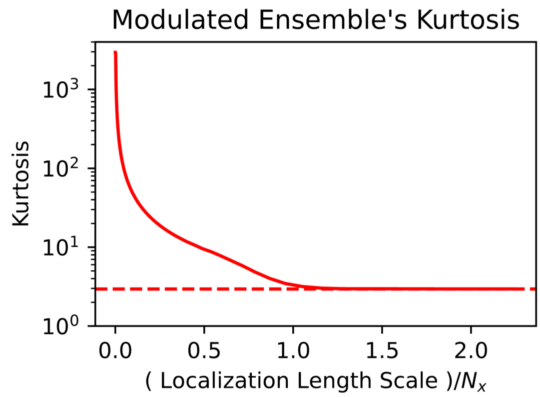

Figure S1. A plot of the modulated ensemble's kurtosis as a function of normalized localization length scale (solid red curve). The kurtosis of the modulated ensemble for infinite normalized localization length scale is indicated by the dashed red line.

where $\left\langle \left(x'_g\right)^p \right\rangle$ is the p-th central moment of state vector element g in the forecast ensemble.

The non-Gaussian characteristics of the modulated ensemble perturbations can be inferred from kurtosis ($Kurt$), which is

$$Kurt\left(v'_g\right) \equiv \frac{\left\langle \left(v'_g\right)^4 \right\rangle}{\left\langle \left(v'_g\right)^2 \right\rangle^2} \approx \frac{N_L \left[\sum_{m=1}^{N_L} \left(\ell_{m,g}\right)^4 \right] \left\langle \left(x'_g\right)^4 \right\rangle}{\left\langle \left(v'_g\right)^2 \right\rangle^2}. \tag{S6}$$

Since the modulated ensemble's variance is identical to the original ensemble's variance,

$$Kurt\left(v'_g\right) \approx N_L \left[\sum_{m=1}^{N_L} \left(\ell_{m,g}\right)^4 \right] \frac{\left\langle \left(x'_g\right)^4 \right\rangle}{\left\langle \left(x'_g\right)^2 \right\rangle^2}. \tag{S7}$$

The fraction is simply the kurtosis of the original ensemble. Thus,

$$Kurt\left(v'_g\right) \approx N_L \left[\sum_{m=1}^{N_L} \left(\ell_{m,g}\right)^4 \right] Kurt\left(x'_g\right). \tag{S8}$$

If the forecast ensemble is drawn from a Gaussian distribution (the kurtosis is always 3), for $N_E \gg 1$, $Kurt\left(x'_g\right) \approx 3$. However, Eq. (S8) states that $Kurt\left(v'_g\right)$ is $N_L \left[\sum_{m=1}^{N_L} \left(\ell_{m,g}\right)^4 \right]$ times of the forecast ensemble's kurtosis (≈ 3). This implies the modulated ensemble is likely non-Gaussian.

To illustrate, suppose the localization matrix is simply an $N_x \times N_x$ identity matrix. This means $N_L = N_x$ and $\boldsymbol{L}$ is also an $N_x \times N_x$ identity matrix. Supposing $\delta_{m,g}$ is the

Kronecker-delta, the kurtosis of the modulated ensemble is

$$Kurt\left(v'_g\right) \approx N_x \left[\sum_{m=1}^{N_x} \delta_{m,g}\right] Kurt\left(x'_g\right) = N_x Kurt\left(x'_g\right) = 3N_x. \tag{S9}$$

In other words,

$$Kurt\left(v'_g\right) \gg 3. \tag{S10}$$

Since the kurtosis of a Gaussian distribution is 3, the modulated ensemble has non-Gaussian statistics.

The ensemble modulation method is also explored numerically using a periodic domain with $N_x = 1000$, $N_E = 100$, and a variety of localization matrices (with Gaussian localization functions). The original ensemble members are drawn from a N_x-dimensional Gaussian distribution with zero mean and identity covariance. Every localization matrix has a unique localization length scale (i.e., the "standard deviation" in the Gaussian localization function). In general, $Kurt\left(v'_g\right)$ decreases from $3N_x$ (≈ 3000) towards a value of 3 (i.e., the Gaussian value) as the localization length scale increases (see Fig. S1). These tests demonstrate that for a large range of (commonly used) localization length scales, ensemble modulation turns Gaussian-distributed ensembles into ensembles with non-Gaussian statistics.

References

Bishop, C. H. and Hodyss, D.: Ensemble covariances adaptively localized with ECO-RAP. Part 2: A strategy for the atmosphere, Tellus, Series A: Dynamic Meteorology and Oceanography, 61 A, https://doi.org/10.1111/j.1600-0870.2008.00372.x, 2009.

Bishop, C. H. and Hodyss, D.: Adaptive ensemble covariance localization in ensemble 4D-VAR state estimation, Monthly Weather Review, 139, https://doi.org/10.1175/2010MWR3403.1, 2011.

Bishop, C. H., Whitaker, J. S., and Lei, L.: Gain form of the ensemble transform Kalman Filter and its relevance to satellite data assimilation with model space ensemble covariance localization, Monthly Weather Review, 145, https://doi.org/10.1175/MWR-D-17-0102.1, 2017.