UNIVERSITY OF OKLAHOMA

GRADUATE COLLEGE

REDUCING SAMPLING ERRORS OF THE BACKGROUND ERROR COVARIANCE ESTIMATE IN ENSEMBLE-BASED DATA ASSIMILATION

A DISSERTATION

SUBMITTED TO THE GRADUATE FACULTY

in partial fulfillment of the requirements for the

Degree of

DOCTOR OF PHILOSOPHY

By

BO HUANG
Norman, Oklahoma
2020

REDUCING SAMPLING ERRORS OF THE BACKGROUND ERROR COVARIANCE ESTIMATE IN ENSEMBLE-BASED DATA ASSIMILATION

A DISSERTATION APPROVED FOR THE
SCHOOL OF METEOROLOGY

BY THE COMMITTEE CONSISTING OF

Dr. Xuguang Wang, Chair

Dr. Deepak Devegowda

Dr. Michael Richman

Dr. Louis Wicker

Dr. Steven Cavallo

Acknowledgements

I would like to express my sincere thanks and gratitude to my doctoral advisor, Dr. Xuguang Wang, for her constant passion and patience to support and guide me in my entire doctoral study. Without her inspiration, I never would have made such progress in my research. I also would like to thank my doctoral committee, including Drs. Michael Richman, Louis Wicker, Steven Cavallo and Deepak Devegowda, for attending my regular committee meetings and providing valuable suggestions during my doctoral study. I also have to thank external collaborators who contribute a lot to my doctoral research, including Dr. Craig Bishop from NRL, Drs. Daryl Kleist, Ting Lei, Rahul Mahajan and Fanglin Yang from EMC, Drs. Jeff Whitaker and Lili Lei from ESRL, for their fruitful discussion and technical assistance. In the meantime, I must extend my thanks to the staffs and students from the School of Meteorology who bring lots of convenience and joy to my entire doctoral study.

Computational resources for my dissertation research were provided and supported by the NOAA research and development HPC, the OU Supercomputing Center for Education and Research at the University of Oklahoma, and the NCAR Computational and Information Systems Lab. I really appreciate their constant help with my technical computing issues.

Lastly, I thank all of my family and friends for their encouragement and support on my path to the PhD.

Table of Contents

List of Tables

List of Figures

Abstract

Ensemble-based data assimilation (DA) estimates forecast error in a flow-dependent fashion, in contrast to the traditional variational DA assuming static forecast error. Due to computational constraints, however, accuracy of the ensemble background error covariance estimate is compromised by sampling error. Direct increase of ensemble size contributes to reducing sampling error, but the increased computational cost can be prohibitive especially for operational numerical weather prediction (NWP) applications. On the other hand, the localization is commonly applied to reduce or eliminate distant spurious correlations caused by sampling error. But the localization can cause additional imbalance in the analysis and potentially eliminate some distant realistic signal. Therefore, efficient treatment of sampling error remains a challenge for the ensemble-based DA. This dissertation covers three topics of efficiently reducing sampling error and further improving background error covariance estimate in the ensemble-based DA.

Firstly, a valid-time-shifting (VTS) ensemble method is introduced as cost-efficient means to increase ensemble size. The VTS method, either in form of full ensemble members (VTSM) or ensemble perturbations (VTSP), is implemented and examined in the National Centers for Environmental Prediction (NCEP) Global Forecast System (GFS) hybrid four-dimensional ensemble-variational (4DEnVar) DA system. VTSM and VTSP applying a single shifting time interval (e.g., one, two or three hours) triple the baseline background ensemble size from 80 to 240 in the EnVar update, while the overall cost only increases by 23%-27% depending on the selected shifting time interval. Ten-week cycled DA experiments show that VTSP generally improves global temperature and wind

forecasts to 5 days over running the original 80-member ensemble. VTSM generally degrades global forecasts in the troposphere. Both VTSM and VTSP improve tropical storm track forecasts over running the original 80-member ensemble. In particular, three-hour VTSM even shows more accurate tropical storm track forecasts than directly running a 240-member ensemble. Further diagnostics are conducted to understand the potential cause of their different performances.

Secondly, two localization methods commonly applied in the ensemble Kalman filter (EnKF) are mathematically examined and compared, that is, the B-localization method that performs the localization on the background error covariances, and the R-localization method that inflates the observation error variances. Mathematical demonstration suggests that the B-localized background error covariance matrix shows higher rank than the R-localization method. The B-localization method is further realized in the ensemble transform Kalman filter (ETKF) by modulating and extending the background ensemble perturbations. The B-localized ETKF is termed as the high-rank ETKF (HETKF) to distinguish from the classic R-localized ETKF. Cycled DA experiments using the Lorenz model II show that the HETKF outperforms the R-localized ETKF especially for a small ensemble, which is likely associated with the higher rank from the B-localization method.

Lastly, the simultaneous multi-scale DA capability is developed in the GSI-based global hybrid 4DEnVar system by introducing scale-dependent localization (SDL) with and without considering cross-waveband covariances (i.e., SDL-Cross and SDL-NoCross, respectively). SDL applies a different amount of localization to different scales of

ensemble background error covariances, while performing a single-step simultaneous assimilation of all the available observations. One-month cycled DA experiments using the NCEP Finite-Volume Cubed-Sphere Dynamical Core (FV3)-based GFS model show that both SDL-NoCross and SDL-Cross improve general global forecasts to five days over applying fixed uniform localization once at all scales. By including the cross-waveband covariances, SDL-Cross tends to show more accurate global forecasts and tropical storm track forecasts than SDL-NoCross at longer forecast lead times. Compared to the two-waveband SDL experiments, the three-waveband SDL experiment counterparts applying tighter horizontal localization at medium-scale waveband generally show improved (degraded) global forecasts below (above) 50 hPa, except the outperformance of the three-waveband SDL-Cross experiment over the two-waveband SDL-Cross experiment below 50 hPa only lasts for three days. In addition, the three-waveband SDL-Cross experiment even shows more accurate tropical storm track forecasts than the experiment applying the operational level-dependent localization at shorter forecast lead times. Diagnostics are further performed to understand their different performances.

Chapter 1: Introduction

Multiple portions of this dissertation are direct excerpts in Huang and Wang (2018; © American Meteorological Society) and Huang et al. (2019; © American Meteorological Society). These include parts of Chapters 2, 3 and 4.

1.1 Background

Data assimilation (DA, Ghil and Malanotte-Rizzoli 1991) is an objective statistical process that "optimally" combines observations with short-range forecasts (i.e., the background) to provide the best estimate of the "true" state (i.e., the analysis). The resultant analysis is used to initialize model forecasts in the numerical weather prediction (NWP). By ingesting fresh observation information into model forecasts through a cycled DA mode, it avoids significant departure of the model forecasts from the "true" state that will finally occur in a free model run due to inherent model deficiencies (Tribbia and Baumhefner 2004). In addition to producing initial conditions for model forecasts in NWP, DA has other important applications. For example, it can provide reanalysis for climate studies (Dee et al. 2011; Laloyaux et al. 2016) and optimize observation network design (Bishop et al. 2001; Majumdar et al. 2002). Furthermore, it also contributes to improving the NWP models and enhancing the understanding of atmospheric dynamics and predictability (Kalnay 2003).

In theory, the well-known DA methods are generally derived from two different approaches. In the DA methods based on the least square approach, such as the optimal interpolation method (Lewis et al. 2006) and the Kalman filter (Kalman and Bucy 1961), its goal is to find optimal weights of the observations and the background that minimize the analysis

variances. In addition, the variational DA method (Lorenc et al. 2000) is derived from the Bayesian approach that attempts to find the analysis representing the maximum likelihood of the "true" state. The analysis in the variational DA method is obtained by minimizing a cost function that measures the weighted distances of the analysis from the observations and the background. Under the assumption of Gaussian error distribution, both the least square approach and Bayesian approach in fact solve the same problem and arrive at equivalent solutions (Lorenc 1986).

In the DA methods derived from both the least square approach and Bayesian approach, it is essential to accurately estimate the background and observation error covariances. The background and observation error variances determine the magnitude of corrections made from the observations to the model state variables of interest (Lorenc et al. 2000; Kalnay 2003; Lewis et al. 2006). In addition, the background correlations determine how the observation information spreads to or influences the adjacent model state variables (Evensen 1994; Hamill 2006). For example, in the observation-sparse areas, the shape of the analysis increments (i.e., the analysis minus the background) is completely determined by the background correlation structure. The observation errors can be caused by instrument errors, representative errors and observation operator errors (Kalnay 2003). Systematic observation biases need to be removed before assimilating the observations (Kalnay 2003; Hamill 2006). To simplify the DA solution in the operational NWP applications, the error correlations between different observations are normally assumed to be zero, thus leading to a diagonal observation error covariance matrix (Kalnay 2003; Hamill 2006). The estimate or modeling of the background error covariances is especially difficult and challenging, since the "true" state is unknown and the background error covariance

matrix itself has an extreme large dimension (e.g., $O(10^8)$~$O(10^{10})$) in the operational NWP applications (Houtekamer and Zhang 2016). Active research has been conducted for accurate estimate or modeling of the background error covariances.

In the traditional three-dimensional variational (3DVar, Lorenc et al. 2000) DA method, the background error covariances are assumed to be static. It can be estimated in a statistical sense through differences of forecasts at the same valid time (Parrish and Derber 1992), differences between the forecasts and analyses (Fisher 2003), or comparison between the forecasts and observations (Hollingsworth and Lönnberg 1986). For instance, the static background error covariances in the global DA system at the National Centers for Environmental Prediction (NCEP) are calculated from differences of the 48- and 24-hour forecasts valid at the same time. It is also known as the "NMC" (which stands for the National Meteorological Center) method (Parrish and Derber 1992). As a temporal extension of the 3DVar, the four-dimensional variational (4DVar, Rawlins et al. 2007) DA method has the capability of implicitly evolving the background error covariances during the DA window through the linearized model and its adjoint. However, the development and maintenance of the linearized model and its adjoint in the 4DVar is computationally very expensive and challenging especially for the operational large-dimension NWP systems. Overall, utilization of the static background error covariances in the 3DVar and 4DVar, due to lack of considering error spatial/temporal variation (e.g., the flow-dependency), has limited the accuracy of resultant analysis.

To simulate flow-dependent background error, the ensemble Kalman filter (EnKF, Evensen 1994) was proposed by applying the Monte-Carlo approximation in the traditional Kalman filter (Kalman and Bucy 1961). In the EnKF, the background error covariances are

estimated from an ensemble of short-range forecasts. The EnKF is thus able to more realistically estimate spatial, temporal and multivariate error covariances in contrast to the traditional variational DA method assuming static background error covariances. To facilitate practical implementation, different EnKF variants have been developed and will be briefly introduced in Chapter 2. Improved analyses and forecasts from the EnKF have been demonstrated over the traditional variational DA in the regional or global operational DA systems with assimilating simulated or real observations (Houtekamer and Mitchell 1998, 2005; Buehner et al. 2010; Wang et al. 2013; Lorenc 2003a; Zhang et al. 2011). However, due to limited computational resources, the EnKF suffers from sampling error caused by running a small ensemble (more details will be discussed later in the introduction). A comprehensive review regarding the EnKF was presented in Houtekamer and Zhang (2016).

In the most recent, the hybrid DA method was developed by blending the static and ensemble background error covariances to supplement the flow-dependent information of forecast error within a variational framework. In theory, the hybridation of both components of background error covariances can be realized through a direct weighted summation (Hamill and Snyder 2000) or the extended control variable method (Lorenc 2003a). For instance, the hybrid 3D and 4D ensemble variational (i.e., 3DEnVar and 4DEnVar, respectively) systems at NCEP were developed based on the extended control variable method (Wang 2010, Wang et al. 2013; Wang and Lei 2014; Kleist and Ide 2015a,b). Wang (2007) demonstrated that both methods of combining the static and ensemble covariances were mathematically equivalent. In addition, another well-known hybrid DA system is the hybrid-4DVar DA system operationally at the European Center for Medium-range Weather Forecasts (ECMWF) (Bonavita et al. 2016). In the

hybrid-4DVar, the ensemble background covariances are combined with the static background error covariances at the beginning of the DA window. Both the static and ensemble background error covariances are then implicitly evolved through the linearized model and its adjoint in the hybrid-4DVar. This differs from the hybrid 4DEnVar that basically uses the 3DVar framework without requiring the linearized model and its ajdoint. The hybrid 4DEnVar is thus computationally much cheaper than the hybrid-4DVar. Potential advantages of the hybrid DA over the stand-alone variational method or the EnKF were discussed in Wang (2010). By incorporating the ensemble covariances into the variational framework, the hybrid DA is able to consider the flow-dependency of forecast error in contrast to the traditional variational DA method. On the other hand, compared to the EnKF, the hybrid DA is able to alleviate the rank-deficiency issue by including the full-rank static background error covariances, and facilitate applying the dynamical and physical constraints on the analysis within a variational framework as well. The hybrid DA has been applied in several operational NWP centers, such as NCEP, ECMWF, the Canadian Meteorological Center (CMC) and the Japan Meteorological Agency (JMA). The superiority of the hybrid DA over the pure variational DA and the EnKF has been demonstrated with respect to the analyses and subsequent forecasts (Wang et al. 2007a, 2008a,b, 2009, 2013; Wang 2011; Zhang and Zhang 2012; Buehner et al. 2013; Clayton et al. 2013; Kuhl et al. 2013; Gustafsson et al. 2014; Lorenc et al. 2015; Kleist and Ide 2015a; Buehner et al. 2015).

In the ensemble-based DA methods, such as the EnKF and hybrid DA, the ensemble background error covariances are generated by an ensemble of short-range forecasts. This further incurs significant cost increase in the ensemble-based DA method. As discussed earlier regarding the EnKF, the current computational resource is still far from adequate to run an ensemble of a

size comparable with the degrees of freedom of the operational NWP models (Houtekamer and Zhang 2016, Table 1). As a result, sampling error is inevitable in the ensemble-based background error covariances and further degrades the resultant analysis accuracy. The typical feature of sampling error is the spurious correlations at distant regions. If sampling error is not properly treated, it will cause noisy analysis increments and even filter divergence (Hamill 2006). Ideally, direct increase of ensemble size can efficiently reduce sampling error and improve the analyses and subsequent forecasts (Lei and Whitaker 2017; Huang and Wang 2018a). However, the increased computational cost can be prohibitive especially for the complicated operational NWP models. Alternatively, the covariance localization, as an ad-hoc means, is commonly applied to reduce sampling error (Houtekamer and Mitchell 1998, 2001). Its general idea is to gradually reduce or remove the correlations between two distant variables that are assumed to be physically small or spurious. However, the covariance localization can cause some additional issues. For example, it could potentially eliminate the distant but realistic correlations that would be simulated by a very large ensemble (Miyoshi et al. 2014). Meanwhile, it could exaggerate the imbalance issue in the resultant analysis especially when strong localization is applied (Greybush et al. 2011; Holland and Wang 2013). In addition, some localization methods applied on the observation space may not be appropriate for the assimilation of integral-type observations (e.g., satellite radiances and radar reflectivity) whose physical location cannot be explicitly defined (Campbell et al. 2010). Furthermore, seeking the optimal localization length is very important for the resultant analysis accuracy. But it is not trivial to find the optimal localization length, since it varies with model resolution, ensemble size, observation density, etc. (Ying et al. 2018; Huang et al. 2019). In the practical applications, extensive tuning tests are generally performed to find the

optimal localization length that, for example, minimizes analysis errors. In the recent, advanced localization methods have been developed to account for the spatial-, scale- and temporal-dependency (Bishop and Hodyss 2009c; Anderson 2007; Buehner and Shlyaeva 2015; Buehner 2012a). Some localization methods will be briefly reviewed in Chapter 2.

1.2 Motivation and dissertation overview

Due to limited computational resources, sampling error will remain a major source of error in the ensemble-based DA. Extensive research is needed to properly deal with sampling error. Otherwise, the accuracy of the ensemble-estimated background error convariances will be inevitably compromised by sampling error, thus further degrading the performance of the ensemble-based DA system. Three research topics are covered within this dissertation to efficiently reduce sampling error and further improve the background error covariance estimate in the ensemble-based DA.

The first topic seeks to reduce sampling error by increasing ensemble size in a cost-efficient fashion. A valid-time-shifting (VTS) ensemble method is introduced where the ensemble forecasts at the same initialization time but different valid times are used to populate the background ensemble. Compared to directly increasing ensemble size, VTS is computationally more efficient by taking advantage of the available ensemble forecasts. Specifically, two variants of VTS are designed by utilizing either full shifted ensemble members or shifted ensemble perturbations (referred to as VTSM and VTSP, respectively). Both VTSM and VTSP were implemented and examined in the gridpoint statistical interpolation (GSI)-based hybrid 4DEnVar system (Wang and Lei 2014; Kleist and Ide 2015a). The performance of VTSM

and VTSP is evaluated for the general global forecasts and tropical cyclone track forecasts in the GFS that houses different scales of weather phenomena. In-depth diagnostics are further discussed to understand their performance difference.

The second topic contributes to improving mathematical understanding of two localization methods that are commonly applied to reduce sampling error in the EnKF. One is the B-localization method through a Schur product between a localization matrix and the full background error covariance matrix (Houtekamer and Mitchell 2001). The other one is the R-localization method that inflates the observation error variances farther away from the model grid point of interest (Hunt et al. 2007). The mathematical demonstration suggests that the B-localized background error covariance matrix shows higher rank than that from the R-localization method. Furthermore, it explores a means of realizing the B-localization method within the ensemble transform Kalman filter (ETKF, Bishop et al. 2001) by modulating and extending the raw ensemble perturbations. To distinguish from the classic R-localized ETKF, the B-localized ETKF is termed as the high-rank ETKF (HETKF) due to the higher-rank of the B-localization method. This further motivates a performance comparison of the B- and R-localization methods within the same ETKF algorithm by using the Lorenz model II (Lorenz 2005).

The last topic aims to develop the simultaneous multi-scale DA capability in the GSI-based global hybrid 4DEnVar system at the NCEP. As many operational NWP centers are dedicated to developing high-resolution models, the future global model will be able to resolve much wider ranges of scales. This motivates to develop the multi-scale DA capability in the operational GSI-based global hybrid 4DEnVar system at the NCEP. Two variants of scale-

dependent localization (SDL) (Buehner 2012; Buehner and Shalyeva 2015), with and without considering cross-waveband covariances (referred to as SDL-Cross and SDL-NoCross), are introduced to more efficiently alleviate sampling error. In SDL, different amount of localization is applied to different scales of ensemble background covariances, while performing a single-step simultaneous assimilation of all the available observations. This differs from the sequential or multi-step DA update in Zhang et al. (2009) and Miyoshi and Kondo (2013). The performance of SDL-NoCross and SDL-Cross is first compared with applying fixed uniform localization once at all scales for the general global forecasts and tropical cyclone track forecasts in the FV3-based GFS that became operational at the NCEP in the most recent. Furthermore, the performance of SDL-Cross and SDL-NoCross with adopting two and three wavebands is inter-compared to reveal how SDL performs in response to the number of decomposed wavebands. Additional diagnostics are discussed to understand their performance difference.

The dissertation is organized as follows. Chapter 2 first briefly introduces the ensemble-based DA methods that include the generic EnKF and the GSI-based global hybrid 4DEnVar system at the NCEP, and then reviews several popular localization methods. Researches on the above three topics are presented in Chapters 3, 4 and 5, respectively. Chapter 6 summarizes the dissertation and discusses future plans.

Chapter 2: Ensemble-based Data Assimilation

2.1 Ensemble Kalman filter (EnKF)

The ensemble Kalman filter (EnKF, Evensen 1994) was developed by applying the Monte-Carlo approximation within the traditional Kalman filter (KF, Kalman and Bucy 1961) concept. Originally, the Kalman filter is introduced for the linear model system. Specifically, a linear model is used to advance forward both the model state and background error covariances. The linear observation operator is applied in the calculation of the innovation and Kalman gain in the KF. In order to incorporate some degrees of nonlinearity, the extended KF (EKF, Julier and Uhlmann 1997) was developed with capability of evolving the model state using the nonlinear model and calculating the innovation through the nonlinear observation operator. In the EKF, however, the linearized model and its adjoint are required to evolve the background error covariacnes, and the linearized observation operator is applied in the calculation of the Kalman gain. Because the EKF requires developing the linearized version and its adjoint for the nonlinear model, it is computationally more expensive than the traditional KF. Therefore, the EKF is computationally prohibitive for the operational NWP applications. Furthermore, by introducing the Monte-Carlo approximation within the EKF, the EnKF estimates the background error covariances from an ensemble of nonlinear short-range forecasts without requiring the linearized model and its adjoint. The EnKF is thus computationally more efficient compared to the EKF, and becomes possible for the operational NWP applications.

The EnKF consists of two steps: the forecast step and the DA step. In the forecast step, it runs an ensemble of independent short-range forecasts (i.e. the background forecasts) which are

used to calculate the background error covariances in the DA update. The EnKF update equations in the DA step are described following the notations in Ide et al. (1997). The analysis $\mathbf{x}^a = [x_1^a, x_2^a, ..., x_n^a]^{\mathrm{T}}$ is obtained by correcting the background $\mathbf{x}^b = [x_1^b, x_2^b, ..., x_n^b]^{\mathrm{T}}$ with the observations $\mathbf{y}^o = [y_1^o, y_2^o, ..., y_p^o]^{\mathrm{T}}$ weighted by the Kalman gain $\mathbf{K}$,

$$\mathbf{x}^a = \mathbf{x}^b + \mathbf{K}[\mathbf{y}^o - H(\mathbf{x}^b)], \tag{2.1}$$

and

$$\mathbf{K} = \mathbf{P}^b\mathbf{H}^{\mathrm{T}}(\mathbf{H}\mathbf{P}^b\mathbf{H}^{\mathrm{T}} + \mathbf{R})^{-1}, \tag{2.2}$$

where $\mathbf{x}^a$ and $\mathbf{x}^b$ are the analysis and background vectors with a dimension of $n \times 1$, respectively; $\mathbf{y}^o$ is the observation vector with a dimension of $p \times 1$; H and $\mathbf{H}$ are the nonlinear and linearized observation operators, respectively; $\mathbf{P}^b$ denotes the background error covariance matrix with a dimension of $n \times n$; and $\mathbf{R} = \mathbf{I}r^2$ is the diagonal observation error covariance matrix with a dimension of $p \times p$. For simplicity, all the diagonal elements in $\mathbf{R}$ are set equal to r^2. The Kalman gain matrix $\mathbf{K}$ has a dimension of $n \times p$. The superscripts *a*, *b* and *o* denote the analysis, background and observations, respectively. In the EnKF, $\mathbf{P}^b$ is estimated from a *K*-member ensemble of background forecasts,

$$\mathbf{P}^b = \frac{1}{K-1}\mathbf{X}'^b(\mathbf{X}'^b)^{\mathrm{T}} = \mathbf{Z}^b(\mathbf{Z}^b)^{\mathrm{T}}, \tag{2.3}$$

where K is the ensemble size; $\mathbf{X}'^b = [\mathbf{x}_1'^b, \mathbf{x}_2'^b, ..., \mathbf{x}_K'^b] = [\mathbf{x}_1^b - \overline{\mathbf{x}}^b, \mathbf{x}_2^b - \overline{\mathbf{x}}^b, ..., \mathbf{x}_K^b - \overline{\mathbf{x}}^b]$ is the background ensemble perturbation matrix with a dimension of $n \times K$ and each column

represents the k^{th} ensemble perturbation vector $\mathbf{x}_k'^b$ with a dimension of $n\times 1$; $\overline{\mathbf{x}}^b = \sum_{k=1}^{K}\mathbf{x}_k^b \Big/ K$ is the background ensemble mean vector with a dimension of $n\times 1$; and $\mathbf{Z}^b = \mathbf{X}'^b \big/ \sqrt{K-1}$ is the background ensemble perturbation matrix normalized by a factor of $\sqrt{K-1}$. In the practical application, $\mathbf{P}^b$ is not explicitly calculated due to its very large dimension (e.g., $n\times n$). Instead, the terms associated with $\mathbf{P}^b$ in Eq. (2.2) are calculated as,

$$\mathbf{P}^b\mathbf{H}^{\mathrm{T}} = \frac{1}{K-1}\sum_{k=1}^{K}\left[\mathbf{x}_k^b - \overline{\mathbf{x}}^b\right]\left[H(\mathbf{x}_k^b) - \overline{H(\mathbf{x}^b)}\right]^{\mathrm{T}}, \tag{2.4}$$

and

$$\mathbf{H}\mathbf{P}^b\mathbf{H}^{\mathrm{T}} = \frac{1}{K-1}\sum_{k=1}^{K}\left[H(\mathbf{x}_k^b) - \overline{H(\mathbf{x}^b)}\right]\left[H(\mathbf{x}_k^b) - \overline{H(\mathbf{x}^b)}\right]^{\mathrm{T}}. \tag{2.5}$$

There are two main types of EnKF algorithms: the stochastic EnKF and deterministic EnKF. In the stochastic EnKF, each member assimilates different set of perturbed observations by adding random noises that are consistent with the observation error covariances (Houtekamer and Mitchell 1998, 2001). Statistically, assimilating perturbed observations in the stochastic EnKF ensures that the resultant analysis convariances are consistent with those in the traditional KF. A stochastic EnKF has become operational in the global DA system at the Canadian Meteorological Center (CMC). In contrast, the deterministic EnKF updates the ensemble mean and perturbations separately without requiring perturbing the observations. The deterministic EnKF avoids the sampling error associated with perturbed observations in the stochastic EnKF (Whitaker and Hamill 2002). Specifically, the ensemble mean in the deterministic EnKF is updated by directly assimilating the non-perturbed observations following Eqs. (2.1) and (2.2).

On the other hand, it updates the ensemble perturbations to be consistent analysis covariances as in the traditional KF. To facilitate efficient implementations, different deterministic EnKF variants have been developed, such as the ensemble adjustment Kalman filter (EAKF, Anderson 2001), the ensemble square-root Kaman filer (EnSRF, Whitaker and Hamill 2002), the ensemble transform Kalman filter (ETKF, Bishop et al. 2001) and its local versison (LETKF, Hunt et al. 2007). A recent review of the EnKF and related scientific questions is provided in Houtekamer and Zhang (2016). The operational global EnKF system at the NCEP currently adopts the EnSRF algorithm and plans to replace it with the LETKF algorithm in the near future. In this dissertation, the NCEP global EnKF systems with the EnSRF and LETKF algorithms are used for the research in Chapters 3 and 5, respectively. In addition, the ETKF is further examined for comparison of two localization methods in Chapter 4.

2.2 GSI-based hybrid 4DEnVar system

The GSI-based hybrid 4DEnVar system became operational for the global model at NCEP since May 2016. It is used for the research in Chapters 3 and 5 in this dissertation. The EnVar formulation and implementation in the GSI is based on the extended control variable method to ingest the ensemble background covariances within the variational framework (Wang et al. 2013; Wang and Lei 2014; Kleist and Ide 2015a,b). Mathematically, the extended control variable method is equivalent to linearly combining the static and ensemble background covariances (Wang et al. 2007c).

Following the notations in Wang and Lei (2014), the general 4DEnVar formulation is briefly introduced. In the GSI-based 4DEnVar system, the analysis increment $\mathbf{x}'_t$ at time t in a

DA window is calculated as

$$\mathbf{x}'_t = \mathbf{x}'_1 + \sum_{k=1}^{K} \left[\mathbf{a}_k \circ (\mathbf{x}_k^e)_t \right], \tag{2.6}$$

where $\mathbf{x}'_1$ is the analysis increment vector associated with the static background error covariances. The second term on the right-hand side of Eq. (2.6) gives the analysis increment associated with the ensemble background covariances. Specifically, $\mathbf{a}_k$ is the extended control variable vector corresponding to the k^{th} ensemble member. In the current GSI-based 4DEnVar configuration, the same set of $\mathbf{a}_k$ is applied at different time levels. $(\mathbf{x}_k^e)_t$ denotes the k^{th} ensemble perturbation vector at time t normalized by $(K\text{-}1)^{1/2}$, where K is the ensemble size. The sign "$\circ$" performs a Schur product between two vectors or matrices with the same dimension.

The 4D analysis increments are calculated by minimizing the cost function,

$$\begin{aligned} J(\mathbf{x}'_1, \mathbf{a}) = & \frac{1}{2}\beta_1 (\mathbf{x}'_1)^{\mathrm{T}} \mathbf{B}_1^{-1} (\mathbf{x}'_1) + \frac{1}{2}\beta_2 (\mathbf{a})^{\mathrm{T}} \mathbf{A}^{-1} (\mathbf{a}) \\ & + \frac{1}{2} \sum_{t=1}^{L} (\mathbf{y}_t^{o\prime} - \mathbf{H}_t \mathbf{x}'_t)^{\mathrm{T}} \mathbf{R}_t^{-1} (\mathbf{y}_t^{o\prime} - \mathbf{H}_t \mathbf{x}'_t), \end{aligned} \tag{2.7}$$

On the right-hand side of Eq. (2.7), the first term is the background term associated with the traditional static background error covariances $\mathbf{B}_1$. In the second term, $\mathbf{a}$ is a vector that concatenates K vectors of $\mathbf{a}_k$. $\mathbf{A}$ is a block-diagonal matrix that defines the localization for the ensemble background covariances, and it is further expressed as

$$\mathbf{A} = \begin{bmatrix} \mathbf{L} & & & \\ & \mathbf{L} & & \\ & & \ddots & \\ & & & \mathbf{L} \end{bmatrix} \tag{2.8}$$

Each of the K blocks in Eq. (2.8) contains the same predefined correlation matrix $\mathbf{L}$ with unit diagonal elements (Wang et al. 2008b). In the third term of Eq. (2.7), $\mathbf{y}_t^{o\prime}$, $\mathbf{H}_t$ and $\mathbf{R}_t$ are the observation innovation vector, linearized observation operator matrix and observation covariance matrix at time t, respectively. L is the number of time levels spanning the DA window. In addition, parameters β_1 and β_2 determine the contributions of the static and ensemble background covariances where $(1/\beta_1)+(1/\beta_2)=1$ is required (Wang et al. 2008b).

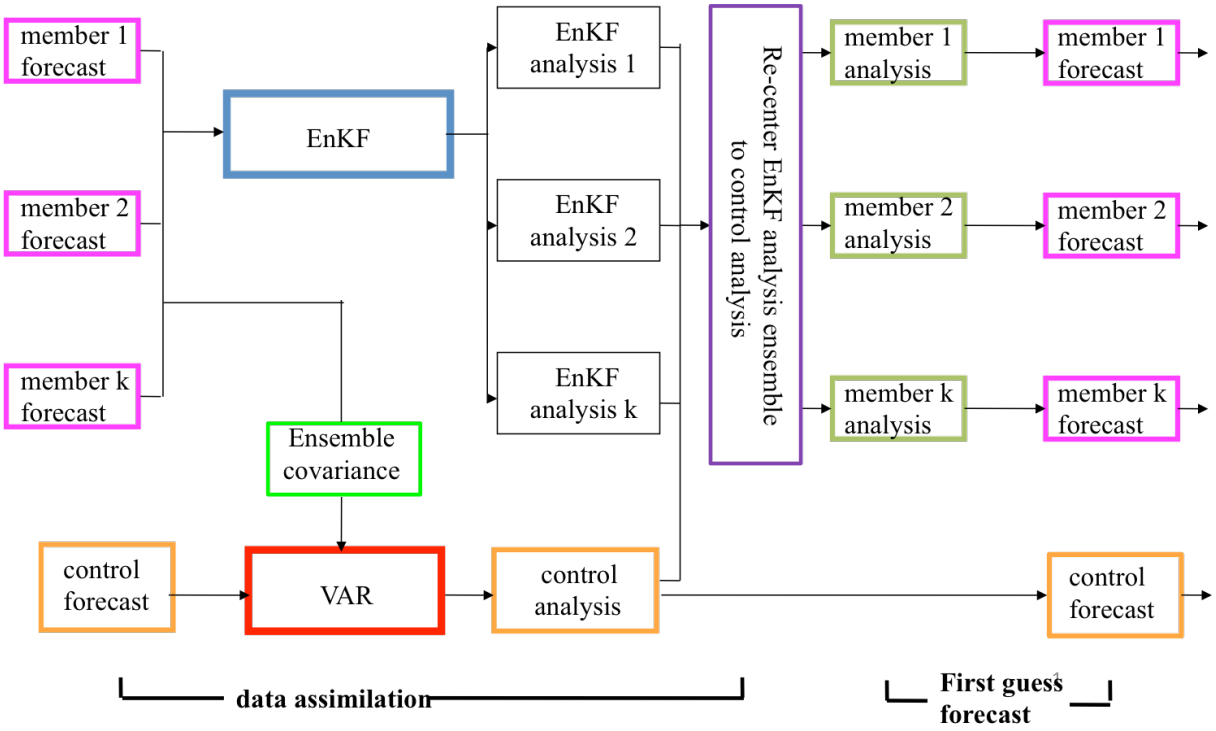

Figure 2.1. Flow chart of the GSI-based global hybrid EnVar system.

Figure 2.1 shows the general flow of a DA cycle in the GSI-based global hybrid EnVar system. It consists of two major components: the variational component and the EnKF component. Specifically, 1) one-member control background forecast in the variational

component and the 80-member ensemble background forecasts in the EnKF component are initialized from the previous DA cycle. The 80-member ensemble is selected in the NCEP GSI-based 4DEnVar system to balance the quality of ensemble-estimated background error covariances and limited computational resources (Kleist and Ide 2015ab). 2) The 80-member ensemble background forecasts are ingested to supplement the flow-dependent ensemble background covariances in the EnVar update of the control background. 3) An EnKF is performed to update the 80-member background ensemble. 4) A re-centering procedure is applied by replacing the 80-member analysis ensemble mean with the control analysis from the EnVar update. 5) The control analysis and re-centered 80-member EnKF analysis are forecasted to next cycle. More details about the operational EnVar configurations and implementations are presented in Chapters 3 and 5 in the dissertation.

2.3 Review of localization methods

As discussed in the introduction, successful application of the ensemble-based DA method relies on covariance localization to alleviate sampling error caused by running a small ensemble (Hamill 2006; Houtekamer and Zhang 2016). Its general idea is to gradually reduce or remove the correlations between two distant variables that are assumed to be physically small or spurious (Houtekamer and Mitchell 1998, 2001). Efficient design of the localization method has been an active topic in the ensemble-based DA research.

First, Houtekamer and Mitchell (1998) applied a cut-off radius to determine the observations to be assimilated at a particular grid point. For example, only the observations within the cut-off radius distance are assimilated to update the variables at that particular grid

point. However, this simple localization form can cause additional noises or imbalance in the resultant analysis due to utilizing an abrupt cut-off radius. Instead, a smoothed localization function was adopted in Houtekamer and Mitchell (2001) that the localization coefficients gradually approach zero at the cut-off radius distance. As a result, the distant observations have less influence in the update of the variables at a particular grid point. In the practical application, the smoothed localization function is commonly defined by the Gaspari-Cohn function that approximates a Gaussian function (Gaspari and Cohn 1999). Such type of distance-dependent localization method can be applied either on the background error covariances (Houtekamer and Mitchell 2005) or the observation error covariances (Hunt et al. 2007) to reduce or remove the correlations between the model variables and distant observations. In the distance-dependent localization that is applied on the observation error covariances, it is required to explicitly specify the physical location of the observations. This requirement makes it inappropriate for the assimilation of the integral-type observations (e.g., satellite radiances and radar reflectivity) whose physical location is difficult to define (Campbell et al. 2010).

To account for the spatial and temporal variations of the localization, some adaptive or dynamic localization methods were developed. In Anderson (2007), the localization function was calculated by minimizing the root-mean-square differences of the regression coefficients computed from multiple groups of ensemble filters. Compared to the fixed distance-dependent localization method, this adaptive localization method improves the accuracy of the observation impact especially at longer forecast lead times (Gasperoni and Wang 2015). In addition, Bishop and Hodyss (2009a,b) proposed calculating the localization function by raising the power of the ensemble correlations. It shows that this adaptive localization method is more advantageous over

the fixed distance-dependent localization method when significant spatial-temporal variation is present in the true correlations. In the recent, Anderson and Lei (2013) developed the empirical localization function (ELF) that minimizes the root-mean-square differences of the posterior ensemble mean and the true values in an Observing System Simulation Experiment (OSSE). In their study, the ELF method produces non-Gaussian localization function for the integral-type observation (e.g., satellite radiances), and outperforms the fixed distance-dependent localization method. Though these adaptive localization methods are more advantageous compared to the fixed distance-dependent localization, it can be very difficult and challenging for operational NWP applications due to significantly increased computational cost.

Furthermore, rather than applying the covariance localization in the physical space, Buehner and Charron (2007) proposed performing the localization in the spectral space. In the spectral localization method, the magnitude of the correlations in the spectral space decreases as the difference between two wavenumbers increases. It is demonstrated that the spectral localization method is equivalent to applying a spatial smoothing of the covariances in the grid space. This feature may contribute to reducing sampling error especially for a small ensemble. This is consistent with the results in Buehner (2012) that the spectral localization method shows more advantages for a small ensemble compared to the spatial localization. In addition, Buehner (2012) proposed a scale-dependent-localization (SDL) method by applying different amount of localization to different ranges of scales of error covariances. In this SDL method, the cross-scale covariances are eliminated. In contrast, another SDL method was developed in Buehner and Shlyaeva (2015) that includes the cross-scale covariances. Compared to applying the fixed uniform localization once at all scales, both SDL methods are demonstrated to improve the

analyses and subsequent forecasts in the operational global and regional NWP applications (Buehner 2012a; Buehner and Shlyaeva 2015; Lorenc 2017a; Caron and Buehner 2018; Caron et al. 2019). Relative performance of these two SDL methods with and without cross-scale covariances can be affected by the accuracy of the ensemble-estimated cross-scale covariances that depends on the ensemble size (Caron et al. 2019). For example, Caron et al. (2019) using a regional 3DEnVar system found that the SDL method without cross-scale covariances tends to perform better than that with the cross-scale covariances for a small ensemble.

To ameliorate the imbalance issue by the localization, Kepert (2009) proposed the variable transformation method where the localization was performed on the streamfucntion-velocity potential rather than the wind components. This method is demonstrated to preserve better balance and further improve the analyses. In Kang et al. (2011), the variable localization method was developed by zeroing out the correlations between two variables that are physically uncorrelated. Its advantage was demonstrated in a carbon cycle data assimilation system.

In this dissertation, two chapters aim to improve the ensemble background covariance estimate by further improving the localization method. In Chapter 4, two aforementioned distance-dependent localization methods are mathematically compared in the generic EnKF context. Chapter 5 describes the development and implementation of the SDL method within the GSI-based global hybrid 4DEnVar system at the NCEP.

Chapter 3: On the Use of Cost-Effective Valid-Time-Shifting (VTS) Method to Increase Ensemble Size in the GFS Hybrid 4DEnVar System

3.1 Introduction

Instead of utilizing the static climatological background error covariances in the traditional variational (Var) systems, the ensemble-based data assimilation (DA) systems are able to simulate the background error covariances in a flow-dependent fashion by using an ensemble of short-range forecasts. One of the best-known forms is the ensemble Kalman filter (EnKF; Evensen 1994). Different variants of EnKF were developed in the last decades for the purpose of efficient implementations (Houtekamer and Mitchell 1998, 2005; Anderson 2001; Bishop et al. 2001; Whitaker and Hamill 2002; Wang and Bishop 2003; Hunt et al. 2007). Recently, the hybrid DA method has shown increasing popularity and been adopted by many operational numerical weather prediction (NWP) centers. The hybrid DA method incorporates the ensemble background error covariances into the Var framework (Hamill and Snyder 2000; Lorenc 2003b; Buehner 2005; Wang et al. 2007c; Wang 2010a). Extensive studies have demonstrated that the hybrid DA method outperforms the stand-alone variational or pure ensemble method (e.g., Wang et al. 2007a, 2008a,b, 2009, 2013; Wang 2011; Zhang and Zhang 2012; Buehner et al. 2013; Clayton et al. 2013; Kuhl et al. 2013; Gustafsson et al. 2014; Wang and Lei 2014; Lorenc et al. 2015; Kleist and Ide 2015a; Kutty and Wang 2015; Buehner et al. 2015).

Within the Monte Carlo approximation, a large-sized ensemble is required to accurately sample the forecast errors in the ensemble-based DA methods. This requirement is especially stringent for high-dimensional NWP models. However, the affordable ensemble size is limited to

be $O(100)$ in many operational centers (Houtekamer and Zhang 2016, Table1) due to the computational constraints. Use of a small-sized ensemble causes sampling errors typically characterized by the remote spurious covariances. If not properly treated, the sampling errors, for example, will incur noisy analysis increments and even cause filter divergence (Hamill 2006). Covariance localization is commonly applied in the ensemble-based DA systems to eliminate the spurious covariances (Houtekamer and Mitchell 2001; Bishop and Hodyss 2009c). However, covariance localization may incur imbalance in the analysis (Buehner and Charron 2007; Wang et al. 2009; Buehner 2012; Holland and Wang 2013) and probably remove the realistic signals in distant regions (Miyoshi et al. 2014). On the other hand, the benefits from directly increasing the ensemble size in the ensemble-based DA systems were demonstrated in the operational or near-operational settings (Bonavita et al. 2012; Houtekamer et al. 2014; Bowler et al. 2017; Lei and Whitaker 2017). Unfortunately, the computational cost is significantly increased due to running a large-sized ensemble.

Instead of directly increasing the ensemble size, early studies explored alternative ways to populate the background ensemble. One method was to include ensemble forecast members valid at the same time but initialized from different previous cycles (Van den Dool and Rukhovets 1994; Lu et al. 2007; Gustafsson et al. 2014). This approach is termed as the time-lagged method hereafter. Such sets of ensemble forecast members with different forecast lengths may be able to sample part of the forecast errors (Van den Dool and Rukhovets 1994). However, previous studies have shown limited success of the time-lagged method to improve the analysis and the subsequent forecasts (e.g. Gustafsson et al. 2014).

In addition, Xu et al. (2008) in a convective-scale forecast context proposed a time-expanded sampling method by taking advantage of the ensemble forecast members that were initialized in the same previous DA cycle but valid at different lead times. This method was inspired to sample the timing and/or phase errors often seen in the convective-scale background forecasts. In its implementation, the ensemble forecast members, valid around but not at the analysis time, were included to populate the background ensemble at the analysis time. Because this method requires shifting the ensemble forecast members valid at different lead times to the analysis time, this method is hereafter denoted as VTSM (Valid-Time-Shifting method for ensemble Members), adapted from the notation used by Gustafsson et al. (2014). The time difference between the analysis time and the valid time of the shifted ensemble forecasts is named as the shifting time interval for brevity. The VTSM method was shown useful for the regional mesoscale EnKF or hybrid-4DVar systems with the assimilation of either simulated or real observations (Xu et al. 2008; Lu et al. 2011; Gustafsson et al. 2014; Zhao et al. 2015).

While the efficacy of the VTSM method has been demonstrated in previous studies for the meso- and convective-scale DA and forecasts in the regional models, its usefulness for a global modeling system featured with a variety of weather phenomena remains to be investigated. For example, the global model houses weather phenomena with different degrees of predictability and scales (e.g., mid-latitude trough and ridge versus tropical storm). The background errors associated with these phenomena can be diverse ranging from timing, phase, magnitude and structure errors and featured with various growth rates. For the weather systems with low predictability, the original background ensemble is likely not able to comprehensively sample the background errors from different sources. In such cases, the populated ensemble by

the VTSM method may have better chance to capture missing sources of background errors. However, as shown in section 3.2, for the more predictable cases, populating the background ensemble by utilizing the ensemble forecast members valid at different lead times may introduce some members that are irrelevant to the background errors. Therefore, the VTSM method is further investigated in a global ensemble-based DA system in this study. This will complement the previous VTSM studies in regional DA applications.

In addition, to ameliorate the potential limitations of the VTSM method while taking advantage of the ensemble forecasts freely available at different lead times to enlarge the ensemble size, a method extended from the VTSM method is explored in this study. Different from the VTSM method, the ensemble members valid at different leading times are re-centered on the original ensemble mean valid at the analysis time. In other words, rather than shifting the ensemble members in the VTSM method, the ensemble perturbations calculated as the deviation from its own ensemble mean are shifted. Hereafter, this method is termed as VTSP (Valid-Time-Shifting method for ensemble Perturbations). In the ECMWF global 4DVar DA system, Bonavita et al. (2016) formed the background covariances by blending the perturbations generated by the VTSP method and those drawn from the climatology. It was found that this blending was beneficial for their global forecasts compared to either using the static climatological background error covariances or the background error covariances estimated from the ensemble perturbations with the same lead time but initialized most recently. In the Met office's global hybrid 4D ensemble-variational system, Lorenc (2017) combined the VTSP approach and the time-lagged approach to further increase the ensemble size. It was found that

the positive impact of the covariances from this populated ensemble was more apparent when proper localization method was implemented.

In this study, the VTSM and VTSP methods were implemented and investigated in the National Centers for Environmental Prediction (NCEP) Global Forecast System (GFS) hybrid four-dimensional ensemble-variational (4DEnVar) DA system (Wang and Lei 2014; Kleist and Ide 2015b). As shown in section 3.2.1 and 3.2.2, the VTSM and VTSP methods produce different impacts in a tropical cyclone example and mid-latitude low example. It is motivated to evaluate and compare the impacts of both VTSM and VTSP methods on the general global forecasts and the tropical cyclone forecasts in a global modeling system. It is expected that the optimal parameters such as the shifting time interval may be dependent on the specific forecast application. Furthermore, various diagnostics are carried out and experiments are designed to reveal the causes of the impacts of the VTSM and VTSP methods. This chapter is organized as follows. Section 3.2 describes and illustrates the VTSM and VTSP methods. Section 3.3 describes the experiment design. Sections 3.4 and 3.5 discuss the results and various diagnostics with respect to the general global forecasts and tropical cyclone track forecasts, respectively. The computational cost from the VTSM and VTSP methods is compared with the baseline GFS 4DEnVar experiment in section 3.6. Section 3.7 presents summaries and discussions.

3.2 VTSM and VTSP methods

In the VTS methods, the ensemble forecasts, either in the form of full ensemble members or ensemble perturbations, which are initialized from the same analyses produced by the previous DA cycle but valid at different lead times, are used to populate the background

ensemble at the current analysis time. Figure 3.1 illustrates the original background ensembles and the VTS-populated background ensemble being ingested into the 4DEnVar variational update within a 6-hour DA window of 0300-0900 UTC. In the original GFS 4DEnVar system (Fig. 3.1a), an ensemble of 80-member forecasts out to the 9-hour lead time is initialized from the analyses produced in the previous DA cycle. Considering producing a 3-hourly temporal resolution of the ensemble within the 6-hour DA window, the analysis increments in the 4DEnVar variational update are produced at three time levels: the beginning (t=0300UTC), middle (t=0600UTC) and end (t=0900UTC) of the DA window, respectively. When applying the VTS methods (Fig. 3.1b), a shifting time interval τ is first selected. Then the background forecasts initialized from the previous DA cycle are shifted both forward and backward for the time length equal to the shifting time interval τ. Specifically, the ensemble forecasts valid at time t-τ and t+τ will be used to supplement the background ensemble at each analysis time t (enclosed by the blue dashed rectangles in Fig. 3.1b). Since these additional ensemble forecasts are not valid but shifted to the time t, they are termed as the shifted ensembles in contrast to the original ensemble valid at time t. As such, for the example given in Fig. 3.1b, the ensemble size is tripled at each time t. The ensemble size can be further enlarged by selecting multiple different shifting time intervals (Lorenc 2017a). For instance, by selecting τ =1, 2 and 3 hours, the populated ensemble size would be seven times as large as the original ensemble. However, these ensemble forecasts with smaller lead time differences can be strongly correlated and therefore not effectively add additional degrees of freedom or rank. Given the focus of this study is to reveal the impact differences of the VTSM and VTSP methods, only one single shifting time interval is selected as described in the next section. The following two subsections describe and

illustrate the VTSM and VTSP methods using both a tropical cyclone and a mid-latitude closed low as examples.

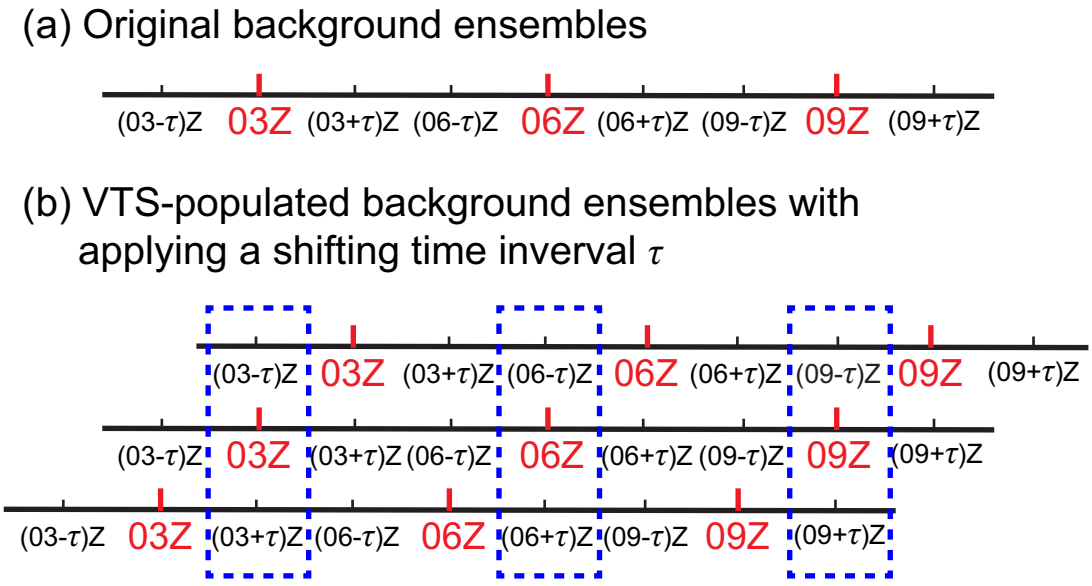

Figure 3.1 Illustration of (a) the original background ensembles and (b) the VTS-populated background ensembles with applying a shifting time interval τ (enclosed by the blue dashed rectangles) being ingested into the 4DEnVar variational update at the three analysis time levels: the beginning (0300UTC), middle (0600UTC) and end (0900UTC) of a 6-hour data assimilation window.

3.2.1 VTSM method

In the VTSM method, the ensemble at time t is populated by directly including the original ensemble forecast members valid at time $t-\tau$ and $t+\tau$. The populated ensemble mean is the average of the original ensemble means valid at time $t-\tau$, t, and $t+\tau$. The populated ensemble perturbations are calculated as the deviation of each member from this populated ensemble mean and are used to compute the ensemble background error covariances. As shown in Eq. (A1) in the appendix A, the background ensemble covariances after applying the VTSM method are calculated by summing up two components. The first component approximately represents an

average of the ensemble covariances from the original and shifted ensembles. The second component is the contribution from the ensemble mean differences between the original and shifted ensembles. As to be discussed in the next subsection, the first component is approximately equivalent to the VTSP-populated ensemble covariances. In the forthcoming examples featuring systems of different scales and predictabilities, the role of the second component from the VTSM-populated ensemble covariances will be discussed.

Figure 3.2 first using typhoon Usage (2013) illustrates that the VTSM method, by directly utilizing the ensemble forecast members at different lead times, can better sample the timing or phase errors, consistent with the early study in Xu et al. (2008). As seen in Fig. 3.2a, a westward location error is observed from the original 6-hour background ensemble mean relative to the verifying ECMWF analysis. The original background ensemble has relatively large spread, indicating the large uncertainty of the forecasts. By shifting the background ensemble forecast members at the 3- and 9-hour lead times to the 6-hour lead time via the VTSM method (Fig. 3.2b), the spread of the ensemble is increased. In the VTSM-populated ensemble, more members enclose the verifying ECMWF analysis especially due to the shifted 3-hour background ensemble members. This result suggests the VTSM-populated ensemble may better sample the location errors than the original 6-hour ensemble. As shown in this typhoon example, the ensemble means at the three different lead times still reside within the envelope of the original 6-hour background ensemble (e.g., the blue curves in Fig. 3.2b), the contribution from the ensemble mean differences is less likely to dominate the VTSM-populated ensemble error covariances as shown in Eq. (A1) in the appendix A.

On the other hand, as discussed in the introduction, the VTSM method may introduce irrelevant members. For example, Fig. 3.2g shows a mid-latitude closed low example where both the phase and structure are accurately predicted by the 6-hour background ensemble mean. The ensemble encloses the verifying ECMWF analysis and the spread of the ensemble is small. The VTSM-populated ensembles with the 3-hour shifting time interval in this example show three distinct clusters in the eastern section of the mid-latitude low (Fig. 3.2h). The added members fall completely outside the envelope of the original ensemble and the verifying ECMWF analysis, therefore producing irrelevant sampling of the background errors. Consistently, the histogram plot for the geopotential height variable sampled from the grid points within the areas of three distinct clusters (enclosed by black dashed rectangle in Fig. 3.2h) shows three peaks (Fig. 3.2k), significantly altering the background distribution of the original ensemble (Fig. 3.2j). As a consequence, the VTSM-populated ensemble violates the Gaussian assumption typically used in the ensemble-based DA methods. In contrast, for the typhoon example, in spite of the increased spread, the VTSM method does not significantly alter the distribution of the original background ensemble by comparing Figs. 3.2d,e. Furthermore, different from the typhoon example as discussed earlier, in the eastern section of the mid-latitude closed low, the ensemble means at the 3- and 9-hour lead times are located totally outside the envelope of the original 6-hour ensemble. Considering the relatively small spread of the original ensembles, the contribution of the ensemble mean differences in the VTSM-populated ensemble covariances could possibly dominate the VTSM-populated ensemble covariances in this mid-latitude closed low example.

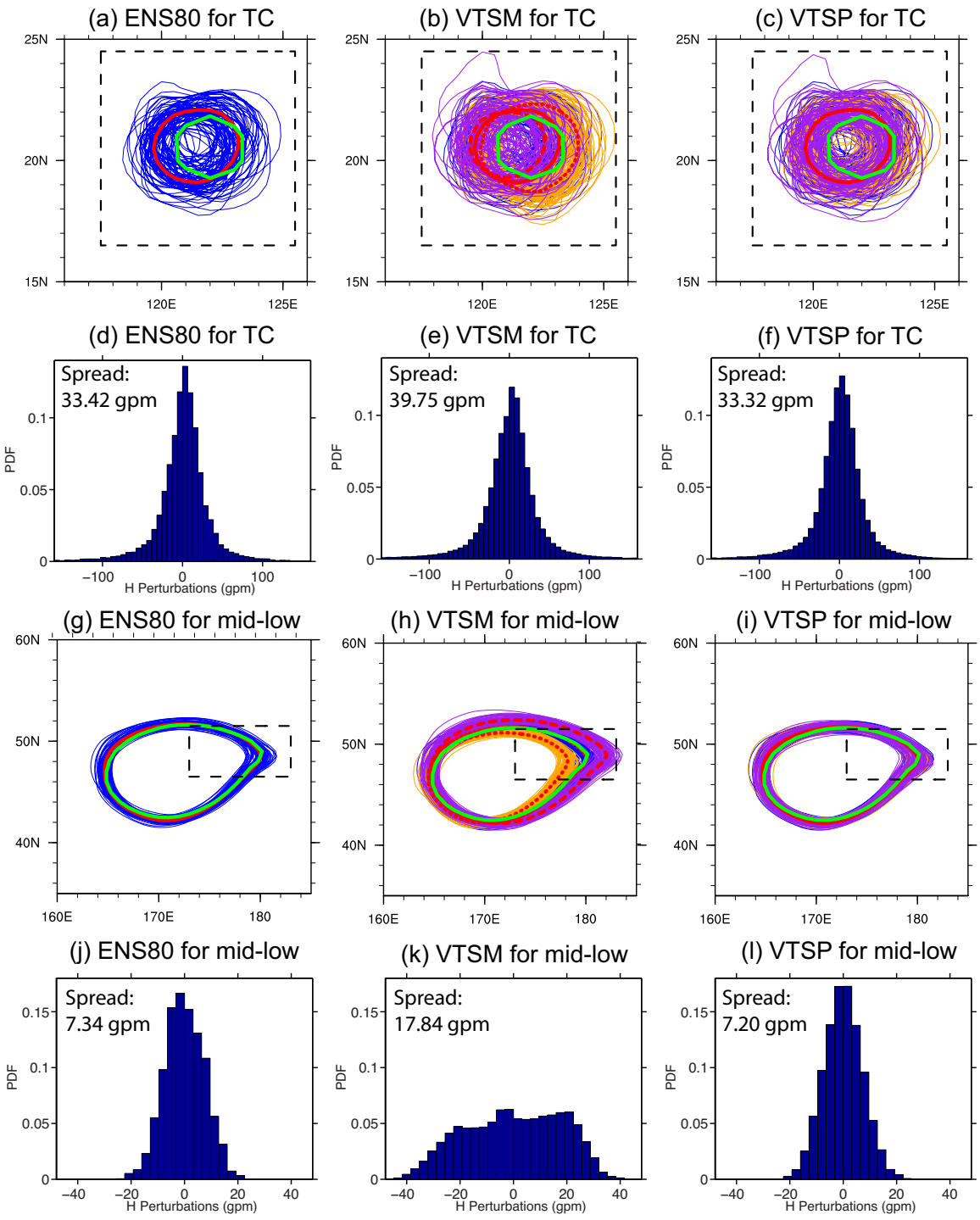

Figure 3.2 Spaghetti-contour plots of (a)-(c) the -120 gpm geopotential height perturbations at 1000 hPa in typhoon Usage (2013) and of (g)-(i) the 1400 gpm geopotential height at 850 hPa in a middle-latitude closed low from the (left) original 80-member background ensemble and (middle) VTSM- and (right) VTSP-populated 240-

member background ensemble with applying a shifting time interval τ = 3 hours at the 6-hour lead time. In (a)-(c) and (g)-(i), the thin blue contours represent the original 6-hour 80-member background ensemble. The thin orange and magenta contours denote the 3- and 9-hour 80-member background ensemble produced by the VTSM and VTSP methods (see texts for the differences), respectively. The thick green contour represents the ECMWF analysis valid at the 6-hour lead time. The short-dashed, solid and long-dashed red contours denote the 80-member background ensemble mean valid at the 3-, 6- and 9-hour lead times, respectively. Histogram plots for (d)-(f) the typhoon Usage (2013) example and (j)-(l) the mid-latitude closed low example by sampling for the geopotential height variable at the grid points enclosed by the corresponding black dashed rectangles in (a)-(c) and (g)-(i) and the averaged spread calculated for the geopotential variables within the corresponding black dashed rectangles is listed on the top left corner correspondingly.

3.2.2 VTSP method

In the VTSP method, the shifted ensemble members at time t are produced by re-centering the original ensemble perturbations at time $t-\tau$ and $t+\tau$ on the original ensemble mean at time t. So the VTSP-populated background ensemble shares the same original background ensemble mean valid at time t. The VTSP method therefore reduces the possibility that the shifted members sample irrelevant background error space. For example, in the mid-latitude low example, different from the VTSM method, the VTSP method produces a populated ensemble without distinct clusters (Fig. 3.2i). Furthermore, the VTSP-populated ensemble (Fig. 3.2l) follows the Gaussian distribution more than the original ensemble (Fig. 3.2j). In the typhoon example, compared to the original ensemble, the VTSP method still increases the chance that the truth is sampled by adding more members enclosing the verifying ECMWF analysis (Fig. 3.2c), although the VTSP method does not increase the spread as much as the VTSM method. In the typhoon example, the VTSP method (Fig. 3.2f), like the VTSM method, does not show significant change of the distribution of the original background ensemble.

As shown in Eq. (A3) in the appendix A, the VTSP method by design functions as averaging the ensemble covariances at three different valid times. Therefore, the VTSP method produces a temporal smoothing effect on the ensemble covariances. In a chaotic system, temporal smoothing has similar effect as spatial smoothing (Raynaud et al. 2008). Buehner and Charron (2007) proved that the spatial smoothing of the ensemble correlations in the grid space was equivalent to applying the localization in the spectral space. Raynaud et al. (2008, 2009) applied a spatial averaging technique on the ensemble background variances to reduce the sampling errors that often have smaller scales compared to the background errors. Therefore, it is expected that the built-in smoothing effect in the VTSP method can contribute to eliminating the spurious small covariances caused by sampling errors. As discussed in the previous subsection, the VTSP-populated ensemble error covariances are approximately equal to the first component in the VTSM-populated ensemble error covariances. By comparing with the VTSP method, it assists in isolating and evaluating the impact of the ensemble mean differences in the VTSM-populated ensemble error covariances.

3.3 Experiment design

The hybrid 4DEnVar system for the GFS model was developed as an extension of the gridpoint statistical interpolation (GSI) 3DEnVar system (Wang 2010a; Wang et al. 2013; Wang and Lei 2014; Kleist and Ide 2015a,b). In contrast to 3DEnVar, 4DEnVar is able to account for the temporal evolution of the background error covariances by utilizing the 4D ensemble forecast errors, therefore enhancing the assimilation of 4D observations within a DA window. The GSI hybrid 4DEnVar system was operationally implemented at NCEP beginning May 2016.

The DA cycling experiments were carried out for a ten-week period from 0000 UTC 25 July to 1800 UTC 30 September 2013. The conventional and satellite observations operationally used in NCEP Global Data Assimilation System (GDAS) were assimilated every 6 hours. The same observation quality control and bias correction for the satellite radiances were used as in the operational GDAS system (Zhu et al. 2014).

Table 3.1 List of DA experiments

Expts	**Four components in a single 4DEnvar DA cycle**			
	EnVar update	**EnKF update**	**Control background forecasts**	**Ensemble background forecasts**
ENS80	Original 80-member ensemble for ensemble error covariances	80 members	One-member 9-hour forecast	80-member 9-hour forecasts
ENS240	Original 240-member ensemble for ensemble error covariances	240 members	One-member 9-hour forecast	240-member 9-hour forecasts
VTSM240Hτ	VTSM-populated 240-member ensemble by applying a shifting time interval τ (τ = 1, 2 and 3 hours) for ensemble error covariances	80 members	One-member 9-hour forecast	80-member (9+τ)-hour forecasts
VTSP240Hτ	VTSP-populated 240-member ensemble by applying a shifting time interval τ (τ = 1, 2 and 3 hours) for ensemble error covariances	80 members	One-member 9-hour forecast	80-member (9+τ)-hour forecasts

The baseline 4DEnVar experiment (ENS80 in Table 3.1) without applying the VTSM and VTSP methods is configured similarly as the operational system except that a reduced resolution is adopted due to the computational constraints. The dual-resolution configuration is applied with a control or deterministic member running at a relatively high resolution of T670

and an 80-member ensemble running at a relatively low resolution of T254. In the DA step, the control background is updated by adopting the 4DEnVar algorithm where the extended control variable method is used to ingest the 4D ensemble perturbations. Detailed mathematical formula and implementation of 4DEnVar in the GSI variational minimization can be found in Wang and Lei (2014). The ensemble members are updated using the EnKF (Whitaker and Hamill 2002; Whitaker et al. 2008). Following the two-way coupling method (Fig. 1b in Wang et al. (2013)), the EnKF ensemble analyses are re-centered on the control 4DEnVar analyses.

In the 4DEnVar update, the weights of 12.5% and 87.5% are applied on the static climatological and ensemble background error covariances, respectively, as in the operational system. Different from utilizing the hourly ensemble perturbations as in the operational system, three-hourly ensemble perturbations are ingested in GSI 4DEnVar in the current experiments due to computational constraints. To treat the sampling errors associated with the ensemble-based covariances, the covariance localization is implemented by a spectral filter transform for the horizontal and the recursive filter for the vertical (see details in Wang et al. 2013). Following Lei and Whitaker (2017), the level-dependent localization length scales in Fig. 3.3 are applied in the horizontal direction, and a single value of 0.5 scale heights (i.e., the natural log of the pressure) in the vertical direction. The horizontal and vertical localization length scales in the 4DEnVar variational update are the *e*-folding scales. To alleviate the imbalance issue in the control analysis, a tangent linear normal mode initialization constraint (TLNMC, Kleist et al. 2009) is applied to the analysis increments during the minimization of GSI 4DEnVar as in Wang et al. (2013), Wang and Lei (2014), and Kleist and Ide (2015a,b).

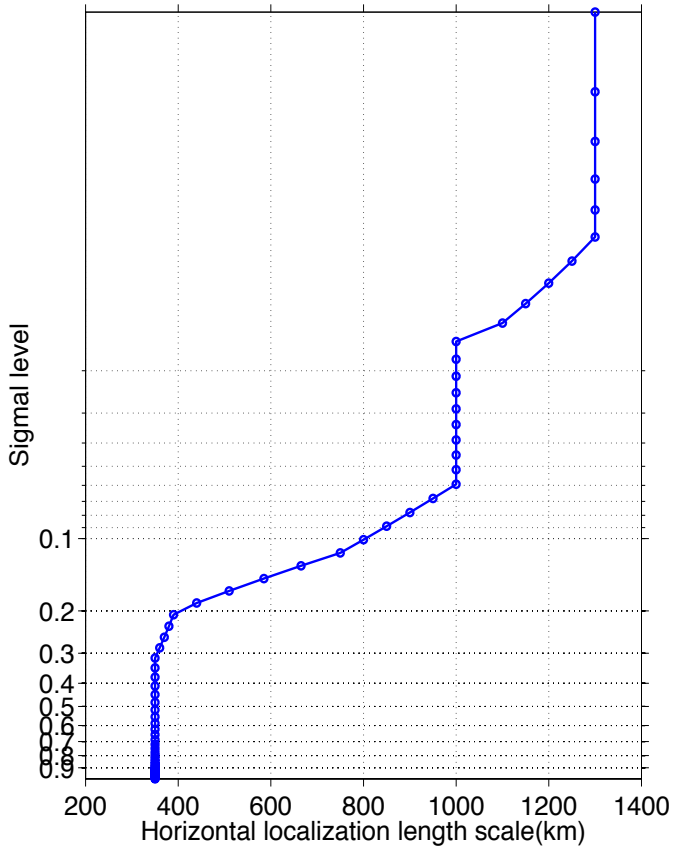

Figure 3.3 Horizontal localization length scales in kilometers as a function of model levels applied in the GSI-based GFS hybrid 4DEnVar system. Note that the horizontal localization length scales are *e*-folding scales.

The 4D serial ensemble square root filter (EnSRF, Whitaker and Hamill 2002) is adopted for the EnKF component of the hybrid system as in the operational system. In the EnKF update, all the observation operators are calculated by GSI. The ensemble background mean at low resolution is used for data selection and quality control so that all the ensemble members are updated by the same set of observations. The same localization parameters used in the 4DEnVar variational update are applied for EnKF. A normalization factor of 0.388 is applied in EnKF to convert the *e*-folding scales to the distance at which the amplitude of Gaspari-Cohn localization function (Gaspari and Cohn 1999) approaches zero. To account for the spread deficiency of the

background ensemble produced by EnKF, the multiplicative inflation (Whitaker et al. 2012) is employed by relaxing the posterior ensemble spread to 85% of the prior ensemble spread. Stochastic parameterization schemes (Palmer et al. 2009; Lei and Whitaker 2016, 2017) are applied to further account for the model uncertainty in the ensemble forecasts.

The GFS model (Environmental Modeling Center, 2020) is configured similarly as the operational system for the control and ensemble forecasts except for running at the reduced horizontal resolutions as discussed before. 64 vertical levels are used. In addition, the 4D incremental analysis update (4DIAU) is applied for both the control and ensemble forecasts instead of the digital filter (DFI; Lynch and Huang 1992) used in the operational GFS model, given the superiority of 4DIAU in suppressing the spurious high-frequency oscillations compared with DFI (Lorenc et al. 2015; Lei and Whitaker 2016, 2017). A 4DIAU implementation is planned for the operational GFS 4DEnVar system (Rahul Mahajan 2017, personal communication).

In addition to the baseline experiment, two sets of experiments, named as VTSM240Hτ and VTSP240Hτ in Table 3.1, are designed where the VTSM and VTSP methods are applied to populate the background ensemble before ingested into the 4DEnVar variational update. VTSM240Hτ and VTSP240Hτ denote experiments where the VTSM and VTSP methods are applied to increase the background ensemble size from 80 to 240 for a given shifting time interval τ. Since the shifted ensembles defined in the VTS methods originate from the ensemble forecasts that are initialized from the same analyses produced in the previous DA cycle, three different shifting time intervals (τ = 1, 2, or 3 hours) are experimented within a 6-hour DA window of 0300Z-0900Z. Note that in these experiments, EnKF and ensemble forecasts still run

with 80 members as in the baseline ENS80 experiment. Only the number of ensemble members ingested to the 4DEnVar variational update is increased from 80 to 240 by using the VTS methods. Finally, the experiment ENS240 is conducted. ENS240 is the same as the baseline ENS80 experiment except that the ensemble size is directly increased from 80 to 240 for each component of the hybrid DA system. As discussed in section 3.6, although 240 members are used in the 4DEnVar variational update for both the ENS240 and VTS experiments, the VTS methods are computationally less costly. ENS240 is therefore used as a reference to reveal to what extent the inexpensive VTS methods can achieve the improvement of or even outperform ENS240. Within similar experiment configurations, Lei and Whitaker (2017) found that the performance of the GFS hybrid 4DEnVar system showed little sensitivity to the localization length scale changes by increasing the ensemble size from 80 to 320. Therefore, our experiments of ENS240, VTSM240Hτ and VTSP240Hτ apply the same localization length scales as the baseline ENS80 experiment. Detailed experiment descriptions are presented in Table 3.1.

3.4 Evaluation of global forecasts

In this section, the performance of the VTSM and VTSP methods on the general global forecasts is evaluated. Various diagnostics are performed to understand the causes of their impacts on the general global forecasts. In section 3.5, the VTSM and VTSP methods are further evaluated on the tropical cyclone track forecasts.

3.4.1 Verification against conventional observations

Root-mean-square-errors (RMSE) of the 6-hour temperature and wind forecasts against the rawinsonde observations were calculated at different pressure levels for all the experiments. Samples were collected from the last eight weeks during the ten-week experiment period to remove the DA spin-up period. The paired *t*-test was performed to examine the significance of the RMSE difference between ENS80 and the other experiments. ENS240 consistently significantly improves the 6-hour temperature and wind forecasts over ENS80 at all pressure levels at or above the 95% confidence level (Figs. 3.4a,d), especially for the global wind forecasts.

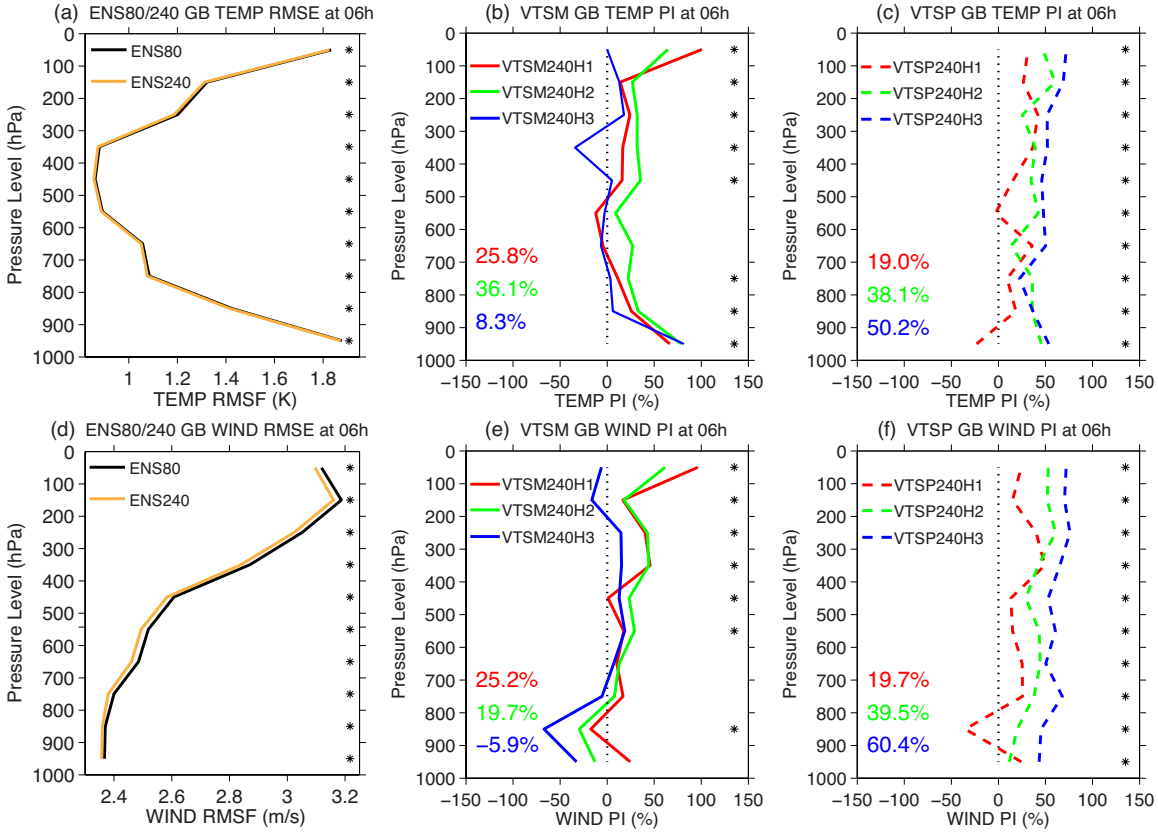

Figure 3.4 Globally and temporally averaged root-mean-square-error (RMSE) of the 6-hour (a) temperature and (d) wind background forecasts in ENS80 (black) and ENS240 (orange) to the rawinsonde observations as a function of pressure levels. Percentage

improvement (PI) relative to ENS240 of the 6-hour (b)(c) temperature and (e)(f) wind background forecasts in (middle) VTSM240Hτ (solid lines) and (right) VTSP240Hτ (dashed lines) experiments with applying a shifting time interval τ = 1 hour (red), 2 hours (green) and 3 hours (blue). The asterisk signs in the left, middle and right panels suggest the RMSE difference from ENS80 in ENS240, VTSM240H2 and VTSP240H3 is significant at or above the 95% confidence level by applying the paired *t*-test, respectively. The percentage number listed in the bottom-left corner of (b)(e) and (c)(f) are the averaged PI over all the pressure levels with the same color indexes applied corresponding to different shifting time intervals in VTSM240Hτ and VTSP240Hτ experiments.

To quantify the RMSE difference of VTSM240Hτ or VTSP240Hτ relative to ENS80 and the extent to which the improvement of ENS240 can be recovered by VTSM240Hτ or VTSP240Hτ, the percentage improvement (PI) of VTSM240Hτ and VTSP240Hτ relative to ENS240 was defined as,

$$\mathrm{PI} = \frac{\mathrm{RMSE(ENS80)} - \mathrm{RMSE(exp)}}{\mathrm{RMSE(ENS80)} - \mathrm{RMSE(ENS240)}} \times 100\% \tag{3.1}$$

where "exp" denotes the experiments of VTSM240Hτ or VTSP240Hτ. Since ENS240 consistently shows smaller RMSE than ENS80, positive PI indicates the improved forecasts of VTSM240Hτ and VTSP240Hτ over ENS80, and vice versa. The averaged PI over all pressure levels was also calculated. VTSP240Hτ with all shifting time intervals shows positive PI for the 6-hour temperature and wind forecasts at almost all pressure levels (Figs. 3.4c,f). The VTSP experiments generally show larger PI with larger shifting time interval. For instance, in terms of averaged PI, compared to VTSP240H1 and VTSP240H2, VTSP240H3 achieves the largest averaged PI of 50.2% and 60.4% for the 6-hour temperature and wind forecasts, respectively. In particular, the improved 6-hour temperature and wind forecasts in VTSP240H3 over ENS80 are statistically significant at or above the 95% confidence level at all pressure levels. On the other

hand, VTSM240H3 shows nearly zero and even negative PI (Figs. 3.4b,e). With a smaller shifting time interval, VTSM240H1 and VTSM240H2 generally show more instances of positive PI. Specifically, in the VTSM experiments, VTSM240H2 (VTSM240H1) shows the best averaged PI of 36.1% (25.2%) for the 6-hour temperature (wind) forecasts. However, this percentage improvement is less than the best-performing VTSP experiment (VTSP240H3). In summary, VTSP240H3 shows the most consistent improvement and therefore recovers the improvement by ENS240 the most for the 6-hour temperature and wind forecasts in all VTS experiments.

3.4.2 Verification against ECMWF analysis

The global forecasts out to 5-day lead times were further verified against the ECMWF analysis with the resolution of 1° × 1° grid (European Centre for Medium-Range Weather Forecasts, 2020). The root-mean-squared-errors (RMSE) between the global forecasts and the ECMWF analyses were calculated every 6 hours and averaged temporally and globally at each pressure level. The paired *t*-test was performed to examine the statistical significance of the RMSE difference from ENS80. Figures 3.5 and 3.6 show the RMSE difference of ENS240 and VTS experiments relative to ENS80 for the global temperature and wind forecasts as a function of forecast lead time and pressure level. ENS240 significantly improves over ENS80 for both the temperature and wind forecasts out to 5 days (Fig. 3.5), which is consistent with Lei and Whitaker (2017).

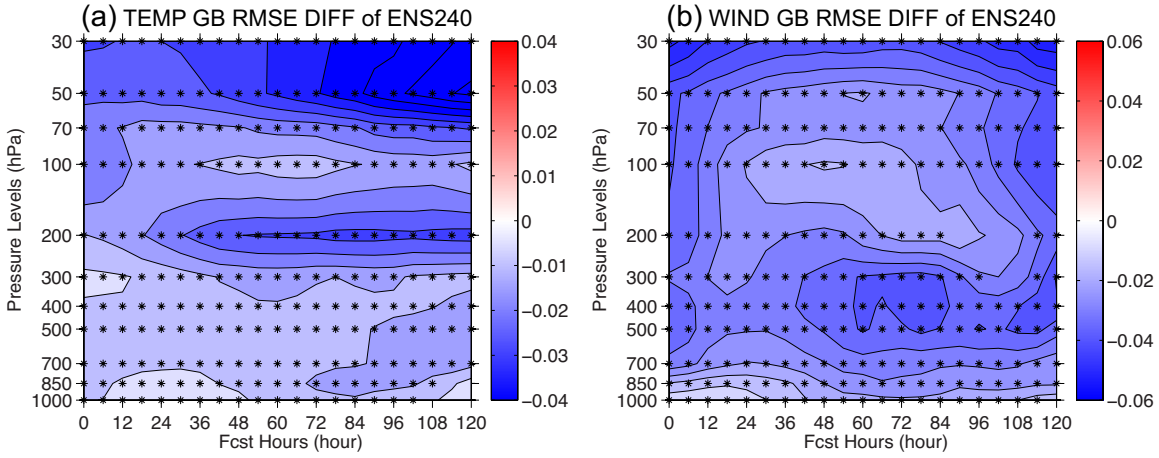

Figure 3.5 Globally and temporally averaged root-mean-square-error (RMSE) difference from ENS80 for the global (left) temperature and (right) wind forecasts in ENS240 against ECMWF analysis as a function of forecast times to 5 days on the horizontal axis and pressure levels on the vertical axis. Blue (red) color indicates smaller (larger) RMSE from the other experiments relative to ENS80. The asterisk signs at the corresponding forecast times and pressure levels indicate that the RMSE difference from ENS80 is significant at or above 95% confidence level by applying the paired *t*-test.

VTSP240Hτ significantly improves temperature and wind forecasts in the stratosphere above 200 hPa over the 5-day lead times. In the troposphere, VTSP240H2 overall is able to maintain the statistically significant improvement out to the 5-day lead time (Figs. 3.6e,k). VTSP240H1 and VTSP240H3 only show significant improvement within the first 3 days and the differences between these two experiments and ENS80 beyond the 3-day lead time is statistically insignificant. VTSM240Hτ consistently improves the global forecasts in the stratosphere above 100 hPa over the 5-day lead times except for VTSM240H3 which shows neutral impacts on the temperature forecasts at early lead times. In the troposphere, however, the VTSM experiments show either nearly neutral or negative impacts. Stronger degradation is found with larger shifting time interval in the VTSM experiments. For example, VTSM240H3 degrades the forecasts below 100 hPa for the entire 5-day period (Figs. 3.6c,i).

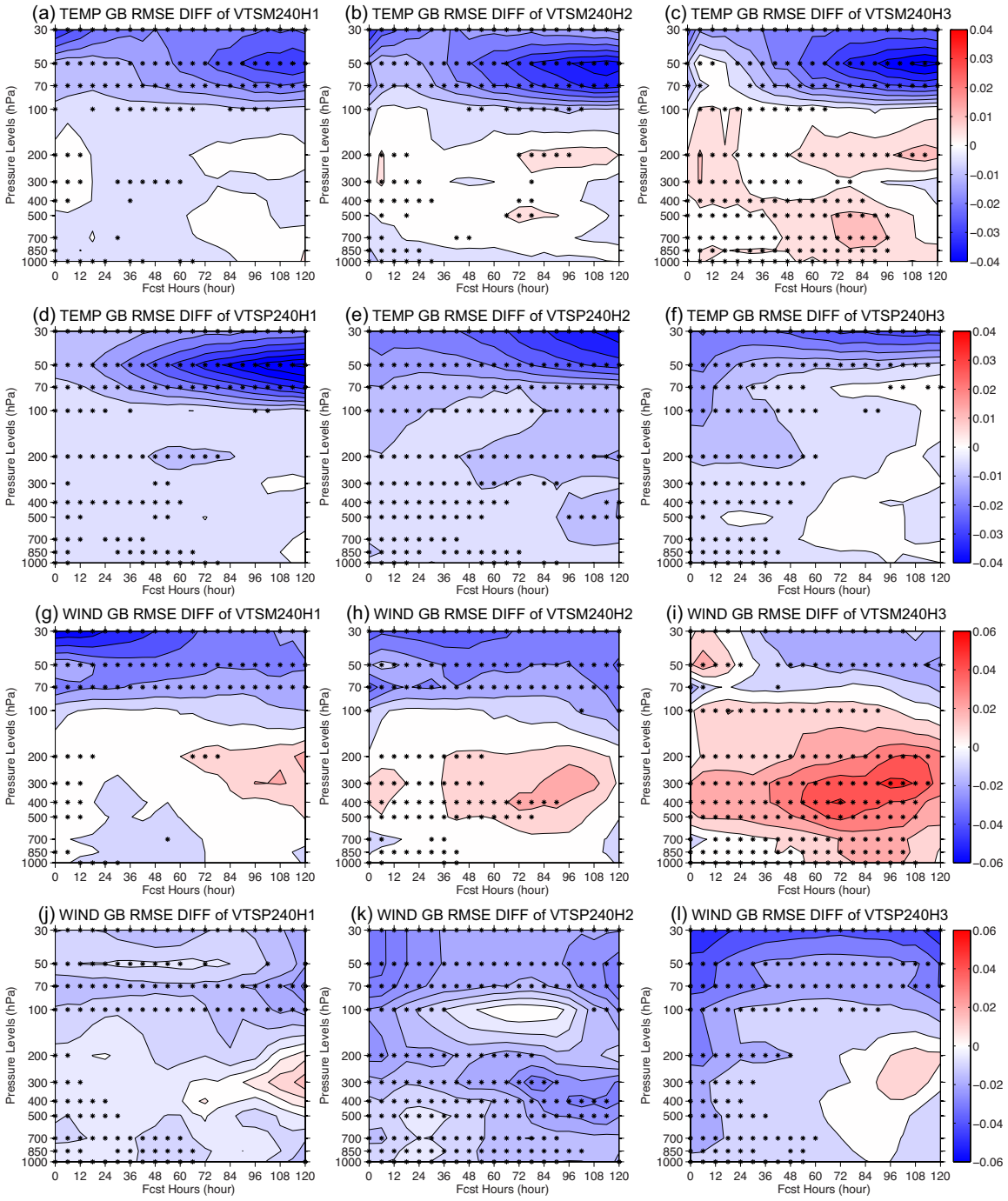

Figure 3.6 As in Fig. 3.5 but for the global (a)-(f) temperature and (g)-(l) wind forecasts against ECMWF analysis in (a)-(c) and (g)-(i) VTSM240Hτ and (d-f) and (j)-(l) VTSP240Hτ experiments with applying a shifting time interval τ = (left) 1 hour, (middle) 2 hours and (right) 3 hours.

In summary, among the experiments of VTSM240Hτ and VTSP240Hτ, VTSP240H2 shows the most consistent improvement for the global temperature and wind forecasts verified against ECMWF analysis. The improvement of VTSP240H2 bears similar structure as that in ENS240 by comparing with Fig. 3.5, though the magnitude is generally smaller. To further understand the causes of the different impacts of the VTSM and VTSP methods on the general global forecasts, in the next several subsections, aspects including ensemble correlation and spread, the effective rank in the ensemble-based covariance matrices are further examined in a global context using ENS240 as a referencing truth.

3.4.3 Accuracy of ensemble correlations

As discussed in the introduction, sampling errors due to the limited ensemble size manifests themselves with spurious correlations in the ensemble-based covariances. Figure 3.7 shows the 2D temperature ensemble auto-correlations at the 6-hour lead time between the central grid point (marked by the black dot) and other grid points for a mid-latitude low case. The ensemble correlations in the referencing ENS240 experiment display flow-dependent structures stretching along the geopotential height contours (Fig. 3.7a). Compared to ENS240, ENS80 shows three spurious negative correlation areas away from the central grid point (Fig. 3.7b) that do not appear in ENS240. VTSP240H3 (Fig. 3.7d) is able to remove the spurious correlations shown in ENS80 and maintain a similar structure as ENS240. These results illustrate the effectiveness of the VTSP method in alleviating the sampling errors, possibly owing to its smoothing effect as discussed in section 3.2.2. On the other hand, VTSM240H3 shows a largely different correlation structure (Fig. 3.7c) compared to ENS240, characterized by the expanded

negative correlation areas away from the centered grid point. As discussed in section 3.2, by comparing with VTSP240H3, the deterioration of the estimated correlations in VTSM240H3 could be caused by the inclusion of the ensemble mean differences that may fail to appropriately sample the forecast errors at the 6-hour lead time (as discussed in Figs. 3.2h,k).

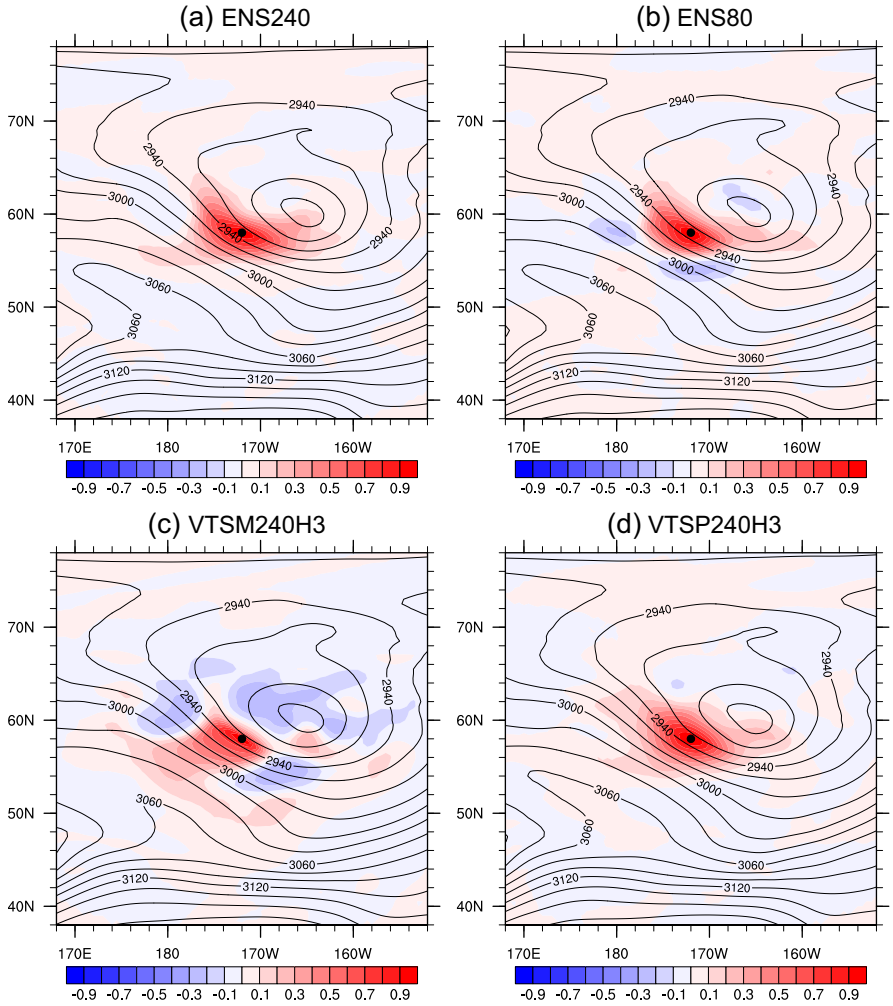

Figure 3.7 700 hPa temperature ensemble auto-correlations (color shaded) between the centered grid point (marked by the black dot) and other grid points calculated from the original 6-hour background ensemble in (a) ENS240 and (b) ENS80, and the (c) VTSM- and (d) VTSP-populated 6-hour background ensembles with applying a shifting time interval $\tau = 3$ hours. The solid black contours represent the geopotential heights of the 6-hour background ensemble mean at 700 hPa.

To quantify the accuracy of the background ensemble correlations, the ensemble correlation samples were collected after the spin-up period. Each correlation sample consists of a 2D box covering an area of 40° × 40° as used in Fig. 3.7. The size of the box is slightly larger than doubling the localization length scale. In each cycle, 165 boxes, evenly distributed over the globe, are selected at 850 hPa, 500 hPa and 200 hPa. Within each box, the auto- and cross-correlations for the temperature and zonal wind variables are calculated between the centered grid point and other grid points. The absolute value of the relative correlation error (ARCE) with ENS240 as the referencing truth is defined as,

$$\mathrm{ARCE(exp)} = \frac{\mathrm{abs[Corr(exp) - Corr(ENS240)]}}{\mathrm{abs[Corr(ENS240)]}} \tag{3.2}$$

where, "exp" denotes the experiments of ENS80, VTSM240Hτ and VTSP240Hτ, "Corr" denotes the background ensemble correlations and "abs" is the absolute sign. Only samples of the absolute correlation values larger than 0.0001 in ENS240 were collected to calculate ARCE. This setting of the threshold intends to reduce the chance of contaminating the averaged ARCE statistics by a limited number of extremely large ARCE produced by a very tiny denominator in Eq. (3.2). To further quantify the error reduction or increase relative to the baseline ENS80 experiment, another metric, ARCE difference (ARCED) is defined as,

$$\mathrm{ARCED(exp)} = \mathrm{ARCE(ENS80)} - \mathrm{ARCE(exp)} \tag{3.3}$$

Positive ARCED suggests the improved correlation accuracy from VTSM240Hτ or VTSP240Hτ relative to ENS80, and vice versa. To evaluate the accuracy of ensemble correlations as a function of the value of the underlying correlations, ten bins with an increasing order of the absolute correlation values are first defined using the absolute correlations in ENS240. ARCE

and ARCED are then grouped and averaged for each bin.

Figure 3.8 shows the ARCE of ENS80 for the temperature variable at 500 hPa in the northern hemisphere (NH), tropical region (TR) and southern hemisphere (SH). Similar results are also found at other pressure levels and for the zonal wind variable (not shown here). In all hemispheres, ARCE decreases as the underlying absolute correlations increase, especially for the small correlations (e.g., the first two bins show sharp decrease of ARCE). In addition, the cross-correlations between the temperature and the zonal wind variables show larger errors than the temperature auto-correlations. The results are consistent with the expectation that for a given ensemble size, it is more difficult to estimate the small correlations and cross-variable correlations using the ensembles.

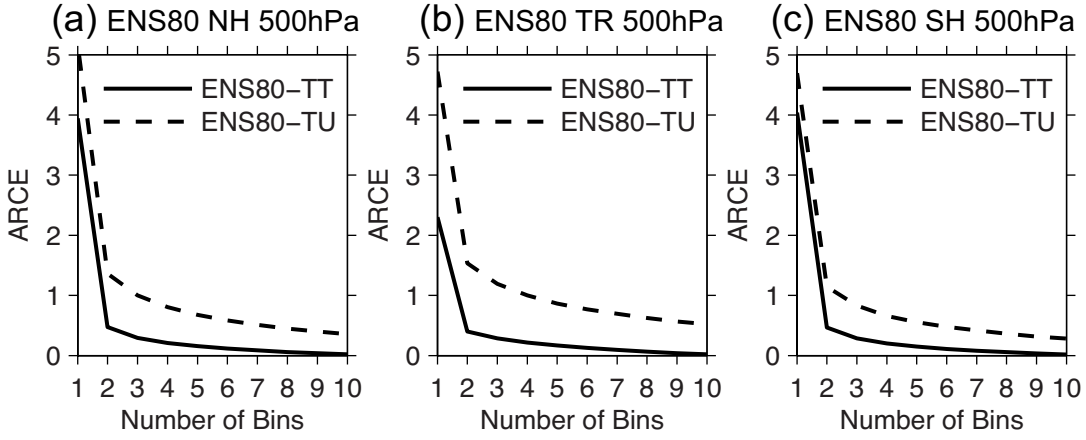

Figure 3.8 Absolute value of relative correlation error (ARCE) of ENS80 at 500 hPa calculated from the 6-hour background ensemble as a function of bin numbers on the horizontal axis in (a) the northern hemisphere (NH), (b) the tropical region (TR) and (c) the southern hemisphere (SH) for the temperature auto-correlations (solid lines, denoted as "TT") and the cross-correlations between the temperature and the zonal wind (dashed lines, denoted as "TU"). Larger bin number on the horizontal axis indicates larger absolute values of the underlying correlations.

Figure 3.9 shows the ARCED of VTSM240Hτ and VTSP240Hτ for the temperature variable at 500 hPa in different hemispheres. Similar results are also found at other pressure

levels and for the zonal wind variable (not shown here). In all hemispheres, VTSP240Hτ consistently improves the correlation accuracy for all bins except for VTSP240H3 which shows slightly degraded cross-correlations in the last three bins in SH. In contrast, VTSM240Hτ degrades the accuracy especially for the small underlying correlations. With the increase of underlying correlations (increasing bin numbers), the improvement in VTSP240Hτ and the degradation in VTSM240Hτ is reduced. Larger improvement in VTSP240Hτ and larger degradation in VTSM240Hτ are generally found for the cross-variable correlations than same-variable auto-correlations. With a larger shifting time interval, VTSP240Hτ generally shows larger improvement and VTSM240Hτ results in larger degradation, respectively. It is speculated that when applying larger shifting time interval in VTSP240Hτ, additional ensemble perturbations added by the VTSP method are more independent from the original background ensemble perturbations, which therefore can more effectively enrich the ensemble. Consistently, VTSP240H3 shows the best percentage improvement for the 6-hour global temperature and wind forecasts (Fig. 3.4c,f). On the other hand, by comparing with VTSP240Hτ, the more severely degraded correlation accuracy with larger shifting time interval in VTSM240Hτ could be attributed to the increased amount of the ensemble mean differences that dominate the VTSM-populated ensemble correlations. It is also noted that VTSM240Hτ shows smallest degradation in TR and largest degradation in SH. Synoptic-scale weather systems typically controlled by the barotropic instability in TR may not evolve as rapidly as those typically controlled by the baroclinic instability in NH or SH (Straus and Paolino 2008). During the experiment period when SH experiences winter time, strong baroclinic instability is expected. As a result, VTSM240Hτ is likely to induce larger amount of ensemble mean differences in SH than that in

TR, thus possibly more severely degrading the VTSM-populated ensemble correlation accuracy in SH.

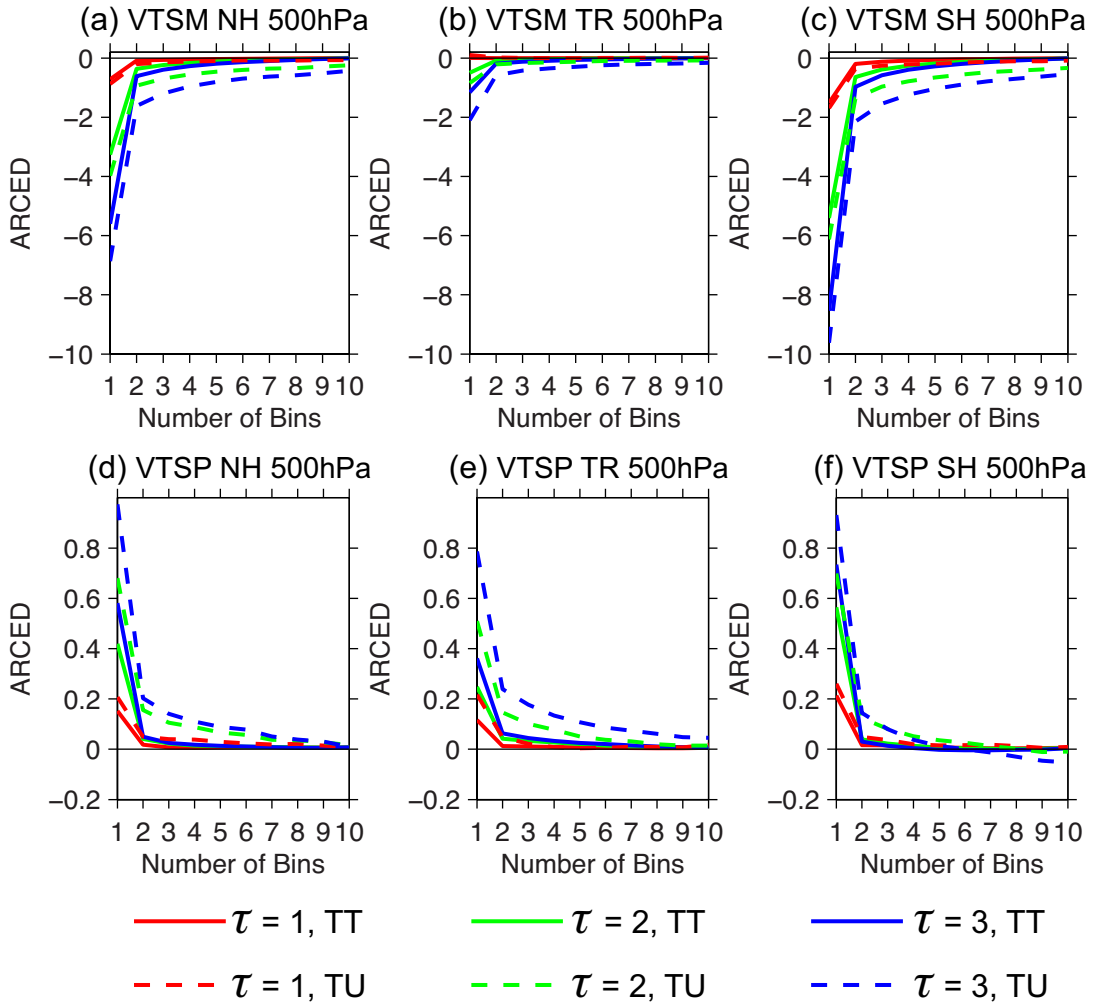

Figure 3.9 Difference of the absolute value of relative correlation error (ARCED) from ENS80 in (top) VTSM and (bottom) VTSP experiments with applying a shifting time interval τ = 1 hour (red), 2 hours (green) and 3 hours (blue) calculated from the 6-hour background ensemble at 500 hPa in (left) the northern hemisphere (NH), (middle) the tropical region (TR) and (right) the southern hemisphere (SH) for the temperature auto-correlations (solid lines, denoted as "TT") and the cross-correlations (dashed lines, denoted as "TU") between the temperature and the zonal wind. The horizontal solid black line represents the ARCED with zero magnitude.

3.4.4 Statistical evaluation of ensemble spread

In this subsection, the relation of the 6-hour background forecast errors and the 6-hour

background ensemble spread are evaluated for all the experiments (Houtekamer and Mitchell 2005; Whitaker et al. 2008; Wang et al. 2013). As shown in Fig. 3.10, in a globally-averaged context, the original background ensemble in ENS80 is under-dispersive for both the 6-hour temperature and wind forecasts in the stratosphere and lower troposphere, but over-dispersive especially for the wind forecasts in the middle troposphere. VTSP240Hτ and ENS240 show negligible spread change from ENS80 (Figs. 3.10c,d), while VTSM240Hτ increases the spread especially with a larger shifting time interval (e.g., VTSM240H3, Figs. 3.10a,b). The increased spread in VTSM240Hτ could be contributed by the ensemble mean differences by comparing with VTSP240Hτ. As a result, VTSM240Hτ is able to alleviate the under-dispersiveness of the original background ensemble in the stratosphere, but exacerbate the over-dispersiveness of the original background ensemble in the middle troposphere. These different effects may explain the improved global temperature and wind forecasts above 100 hPa but the degraded forecasts in the middle troposphere in the VTSM experiments as shown in Fig. 3.6.

In different hemispheres, VTSP240Hτ and ENS240 do not show apparent spread change from ENS80, while the spread increase in VTSM240Hτ in TR is much smaller than that in NH and SH (not shown here). The less spread increase in TR in VTSM240Hτ could be associated with less amount of ensemble mean differences added to the total variance in TR than in NH and SH due to their different types of instabilities as discussed in section 3.4.3.

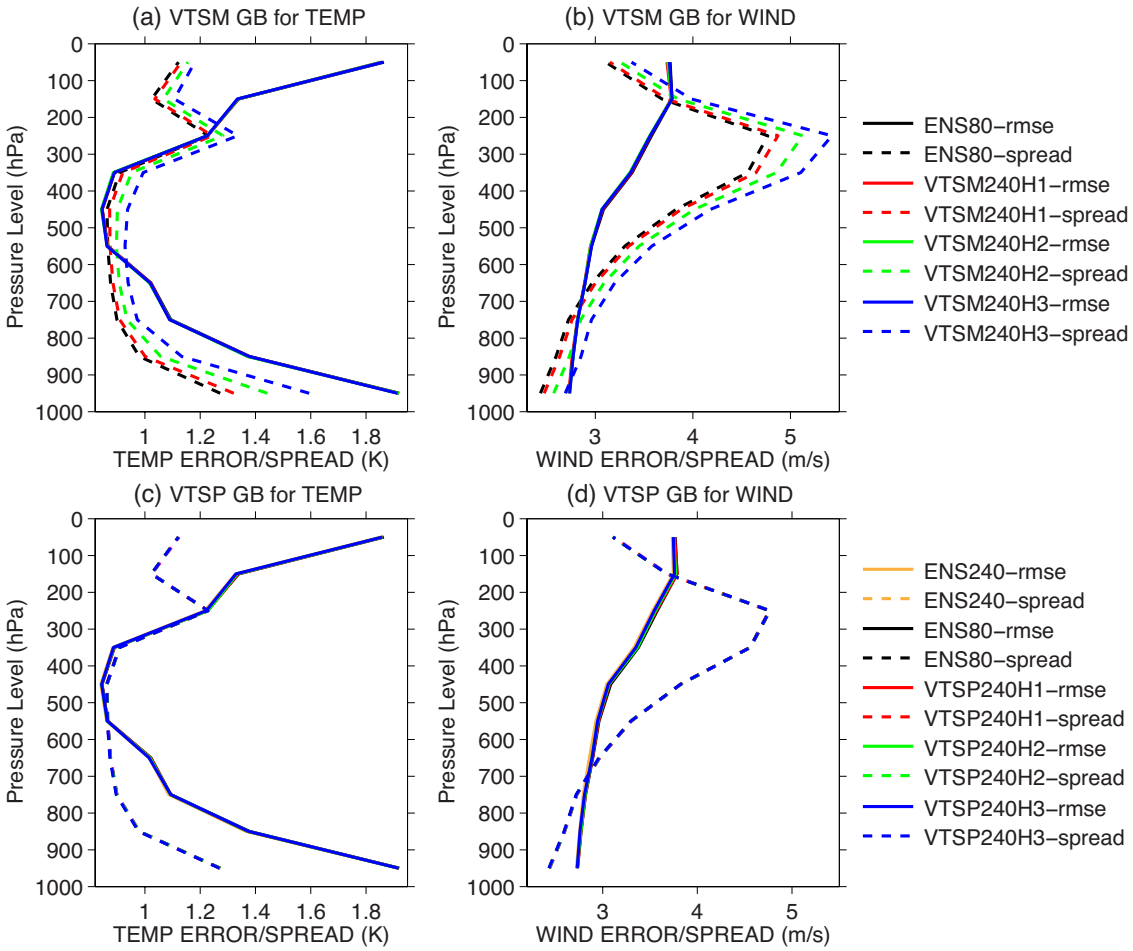

Figure 3.10 Vertical profiles for each experiment of the square root of the globally and temporally averaged innovation variance (solid lines), and the predictions of what it should be if the assimilation assumptions are correct (dashed lines). These predicted values are the square root of the observation error variance plus the variance from the original 6-hour background ensemble in ENS80 (black) and ENS240 (orange) and the (top) VTSM- and (bottom) VTSP-populated 6-hour background ensembles with applying a shifting time interval τ = 1 hour (red), 2 hours (green) and 3 hours (blue) for the (left) temperature and (right) wind forecasts at the 6-hour lead time. Note that many of the curves in (c) and (d) are very similar and have been overplotted by the blue curves which were plotted last.

3.4.5 Measure of effective rank in ensemble covariance matrices

The sampling errors in the ensemble covariances are also manifested in the form of a small number of independent sub-spaces sampled, or sharp eigenvalue spectra of the ensemble

covariances (Wang and Bishop 2003). The E-dimension (Patil et al. 2001; Oczkowski et al. 2005; Kuhl et al. 2007) is therefore calculated to further evaluate the effectiveness of the VTSM and VTSP methods in increasing the effective rank of the ensemble covariance matrix.

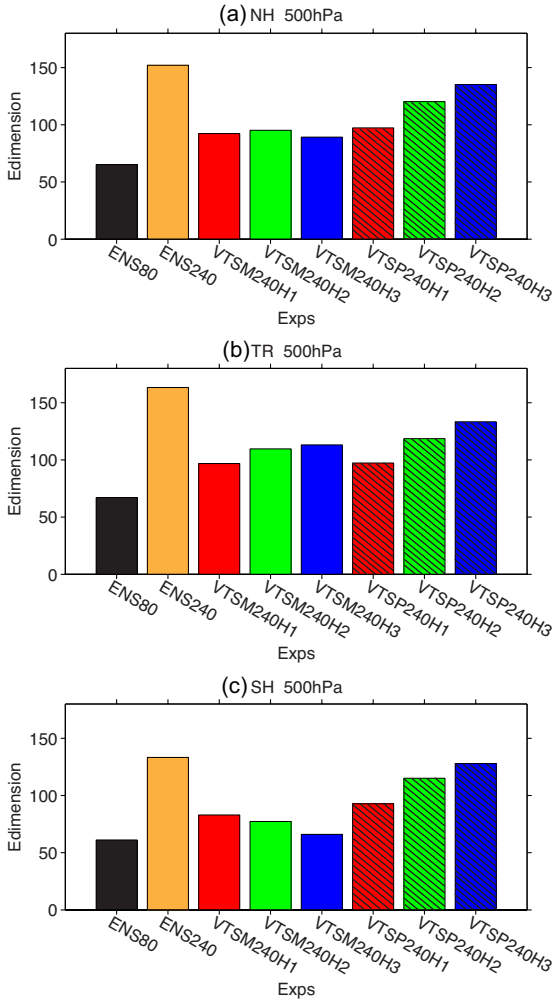

Figure 3.11 E-dimension calculated with using the temperature and zonal wind perturbations at 500 hPa from the original 6-hour background ensemble in ENS80 (black) and ENS240 (orange) and the VTSM- and VTSP-populated (filled with slash lines) 6-hour background ensembles with applying a shifting time interval τ = 1 hour (red), 2 hours (green) and 3 hours (blue) in (a) the northern hemisphere (NH), (b) the tropical region (TR) and (c) the southern hemisphere (SH).

Detailed procedures of calculating the E-dimension were documented in Oczkowski et al. (2005). Specifically, the E-dimension was calculated by collecting the temperature and zonal

wind perturbations at the 6-hour lead time in each box as in section 4c. A total energy rescaling norm is employed following Eq. (26) in Wang and Bishop (2003). The temperature perturbations are multiplied by a factor of $\sqrt{C_p / T_r}$ $\sqrt{\mathrm{C_p/T_r}}$, where, C_p is the specific heat at constant pressure and T_r is the reference temperature with 300K (Palmer et al. 1998). Figure 3.11 shows the E-dimension at 500 hPa in different hemispheres. Although ensemble size is tripled in ENS240, the E-dimension in ENS240 is about 2.3 times as large as that of ENS80 in different hemispheres. VTSM240Hτ and VTSP240Hτ increase E-dimension compared to ENS80. Although having the same background ensemble size of 240, VTSM240Hτ and VTSP240Hτ have smaller E-dimension than ENS240. For a given shifting time interval, VTSP240Hτ shows larger E-dimension than VTSM240Hτ, which suggests that the inclusion of ensemble mean differences in the VTSM-populated ensemble error covariances will reduce the effective rank relative to the VTSP method.

On the other hand, larger shifting time interval results in larger E-dimension in VTSP240Hτ in different hemispheres. This result is consistent with the expectation that separated by larger lead time differences, the ensemble perturbations are more independent. However, with applying a larger shifting time interval in VTSM240Hτ, the E-dimension is decreased in SH while it is increased in TR. In NH, VTSM240H3 also shows slightly decreased E-dimension than VTSM240H1. These results of the VTSM experiments may be attributed to the different amounts of contribution from ensemble mean differences added to the total VTSM-populated ensemble covariances in different hemispheres controlled by different types of instabilities as discussed in sections 4c,d. For example, when a larger shifting time interval is applied in VTSM240Hτ, the relatively larger ensemble mean differences in NH and SH, may

dominate the total VTSM-populated ensemble covariances to a higher degree and thus cause smaller E-dimension as a result. However, in TR, the ensemble mean differences in VTSM240Hτ contributes less to the total covariances and therefore the VTSM-populated ensemble covariances is mostly contributed by the original ensemble perturbations at the three different lead times (e.g., the first component in Eq. (A1)) which are expected to have more degrees of independence with larger time separation.

3.5 Evaluation of tropical cyclone track forecasts

3.5.1 Tropical cyclone track forecast verification

As discussed in sections 3.1 and 3.2, a global forecast system houses diverse weather phenomena. The impacts of the VTSM and VTSP methods can be highly dependent on the scales and predictability of the weather systems of interest. Tropical cyclone (TC) is selected in this section, distinct from the general global forecasts in section 3.4, to further examine the impacts of the VTSM and VTSP methods.

During the experiment period, a total of 25 named storms occurred at the Atlantic and Pacific basins, 12 of which reached the hurricane or typhoon category (Fig. 3.12). NCEP tropical cyclone tracker (Marchok 2002) was used to track the storm locations in the forecasts. The same criteria described in section 4d of Wang and Lei (2014) were used to collect the forecast samples for the purpose of making a homogeneous comparison among different experiments.

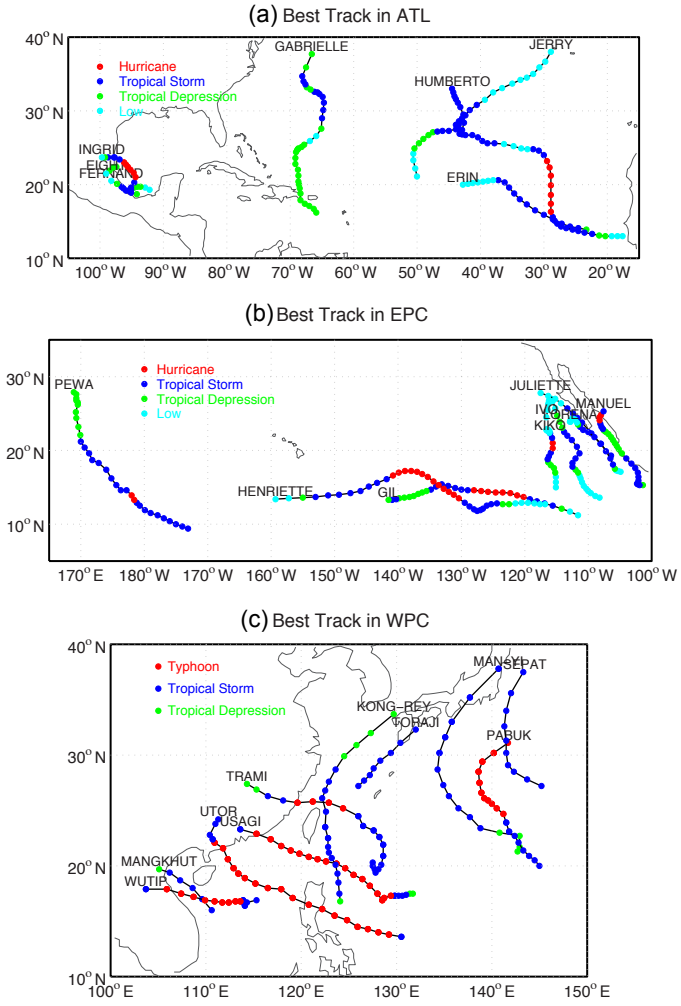

Figure 3.12 Best track of the tropical cyclones during the experiment period in the (a) Atlantic, (b) east Pacific and (c) west Pacific basins.

Figure 3.13a shows the RMSE of the track forecasts verified against the best track data out to 5 days averaged over the 25 storms. Paired *t*-test was conducted to evaluate the significance of the track error difference between ENS80 and the other experiments. ENS240, VTSM240Hτ and VTSP240Hτ all statistically significantly improve the TC track forecasts compared to ENS80 at most lead times out to 5 days.

VTSM240Hτ produces smaller track errors with larger shifting time interval. The best-performing VTSM experiment, VTSM240H3, is statistically significantly better than ENS80

beyond the 1-day lead time. VTSM240H3 performs the best among all VTS and ENS240 experiments. The VTSM experiments show more improvement for the TC track forecasts when applying a larger shifting time interval. This may be expected that the VTSM method with a larger shifting time interval may potentially better sampling the TC timing/phase errors especially when the forecasted TC location at the analysis time experiences large location error. The VTSP experiments do not show strong sensitivity to the shifting time intervals. All VTSP experiments statistically significantly improve the TC track forecasts over ENS80 at most lead times with the 1-hour shifting interval performing slightly better. Although the VTSM and VTSP experiments are less costly than ENS240 (discussed in section 3.6), VTSP240H1 only performs slightly worse than ENS240 within the 4-day lead times and VTSM240H3 even outperforms ENS240 beyond the 2-day lead time.

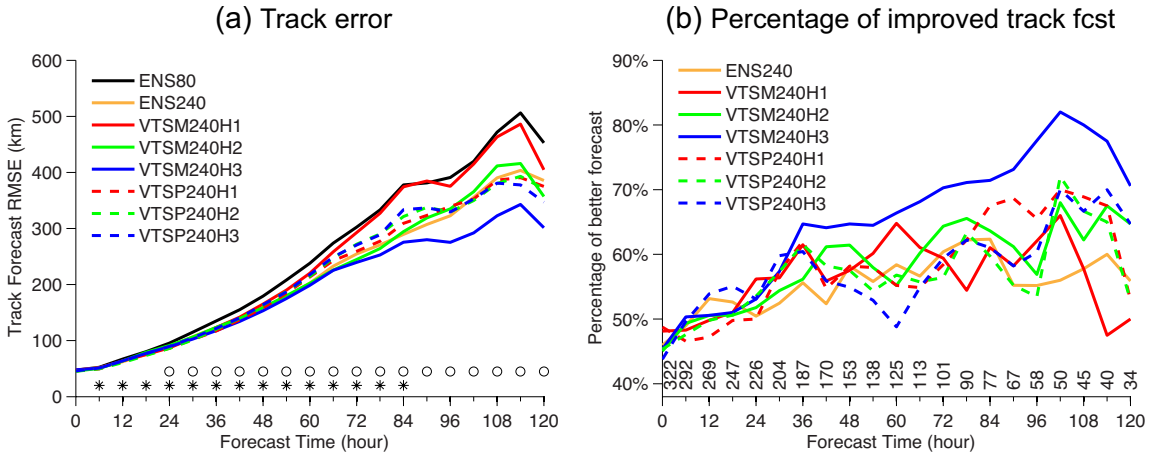

Figure 3.13 (a) Track forecast errors in ENS80 (solid black), ENS240 (solid orange), VTSM240Hτ (solid) and VTSP240Hτ (dashed) experiments with applying a shifting time interval τ = 1 hour (red), 2 hours (green) and 3 hours (blue). The circle (asterisk) signs right above the horizontal axis in (a) indicate that the track error difference from ENS80 in VTSM240H3 (VTSP240H1) is significant at or above the 95% confidence level by applying the paired *t*-test at the corresponding forecast time. (b) Percentage of the track forecasts that are more accurate than that in ENS80 with the same line style and color

indexes applied in (a). The numbers right above the horizontal axis in (b) denote the sample size at the corresponding forecast time.

Following Zapotocny et al. (2008) and Wang and Lei (2014), the percentage of the track forecasts that are more accurate than that in ENS80 was calculated (Fig. 3.13b). Beyond the 1-day lead time, generally more than 50% of the forecasts in each of the experiments of ENS240, VTSM240Hτ and VTSP240Hτ provide more accurate TC track forecasts than that in ENS80 and the percentage is generally increased at longer lead times. Compared to ENS240, VTSP240H1 shows larger percentage of the improved track forecasts at most lead times, especially beyond the 3-day lead time. Consistent with Fig. 13a, VTSM240H3 outperforms ENS240 even more. Specifically, 53.1-82.0% of the track forecasts in VTSM240H3 as opposed to 50.4-62.3% in ENS240 are improved over ENS80 beyond the 1-day lead time. The result that VTSM240H3 even outperforms ENS240 is consistent with the previous studies where the VTSM method was found to improve the meso- and convective-scale weather forecasts where phase and timing errors contribute significantly.

3.5.2 Background track forecast error and ensemble track spread

Given the more accurate track forecasts of VTSM240H3 than ENS240, metrics of ARCE and ARCED with ENS240 as the referencing truth defined in section 3.4.3 are not appropriate to evaluate the ensemble correlation accuracy of VTSM240Hτ and VTSP240Hτ for the TC track forecasts. The E-dimension[1] for the TC track forecasts (not shown here) is similar to the general global forecasts in the TR region (Fig. 3.11b). Briefly, both VTSM240Hτ and VTSP240Hτ

[1] Following section 3.4.5, the samples for calculating the E-dimension for the TC track forecasts were taken by collecting the temperature and zonal wind ensemble perturbations at the 6-hour lead time in a box of 5° × 5° located around the TC center from all 25 tropical storms.

obtain larger E-dimension than ENS80, but smaller than ENS240. The E-dimension in VTSM240Hτ and VTSP240Hτ is further increased with larger shifting time intervals applied. For a given shifting time interval, VTSP240Hτ shows larger E-dimension than VTSM240Hτ.

The improved TC track forecasts of VTSP240Hτ relative to ENS80 is therefore hypothesized to be related to the improved ensemble covariances such as the reduced spurious covariances and increased effective rank as discussed in the general global forecast diagnostics in sections 3.4.3 and 3.4.5. On the other hand, the most accurate TC track forecasts in VTSM240H3 are likely contributed by its capability of capturing background errors from sources that are missing in ENS80, ENS240 and VTSP240Hτ experiments (e.g., model timing or phase errors, as discussed in section 3.2 for such weather systems featured with relatively small scales and low predictability). This capability of the VTSM method in sampling background errors from missing sources is illustrated by the increased spread as discussed in Fig. 3.2b in section 3.2.1.

To further demonstrate this capability of the VTSM method, the background ensemble track spread at the 6-hour lead time is evaluated against the background track error to reveal if the increased spread in VTSM240H3 is another contributor to its most improved TC track forecasts. In Figs. 3.14a-c, the scatterplots were created by collecting a total of 290 paired samples of the background ensemble track spread and absolute background track error from all 25 storms in ENS80, ENS240 and VTSM240H3, respectively. Although the background track errors have similar ranges for all three experiments, VTSM240H3 overall displays a wider range of background ensemble track spread compared to ENS80 and ENS240. Following Wang and Bishop (2003), the spread-skill relationship for the 6-hour background forecast in each

experiment is further evaluated. Given the relatively small number of samples, instead of following Wang and Bishop (2003) which divided the samples into multiple equally-populated bins, two equally-populated bins representing the samples with small and large background ensemble track spread in each experiment (denoted by the blue and red circles in Figs. 3.14a-c, respectively) were formed. In each group, a rank histogram plot was further created. Given the "distance" is evaluated, the rank histogram is formed slightly differently from the traditional scalar rank histogram (Hamill 2001). Specifically, the ranks are formed by collecting the distances between the predicted background ensemble mean storm location and the storm locations predicted from the individual background ensemble members, which is positive definite. Then the number of samples for each rank is determined by throwing the corresponding samples of the distance between the predicted storm location from the background ensemble mean and the observed storm location. Different from the traditional rank histogram plot introduced for the scalar variable in Hamill (2001), a left (right) tail suggests the over-dispersiveness (under-dispersiveness) of an ensemble. For the first group representing small background ensemble track spread, compared to the severe under-dispersiveness of the ensemble in ENS80 and ENS240 (e.g., the right-tailed distribution in Figs. 3.14d,e), the reliability of the ensemble in VTSM240H3 is improved evidenced by a relatively flat distribution (Fig. 3.14f). On the other hand, for the second group featuring with large background ensemble track spread, compared to ENS80 (Fig, 3.14g), the ensembles in ENS240 and especially VTSM240H3 show apparent over-dispersiveness (Figs. 3.14h,i). Therefore, we speculate that the improved reliability of the ensemble in VTSM240H3 for the cases with small background ensemble track spread may be another contributor to its overall outperformance over ENS80 and ENS240 in

terms of the TC track forecasts.

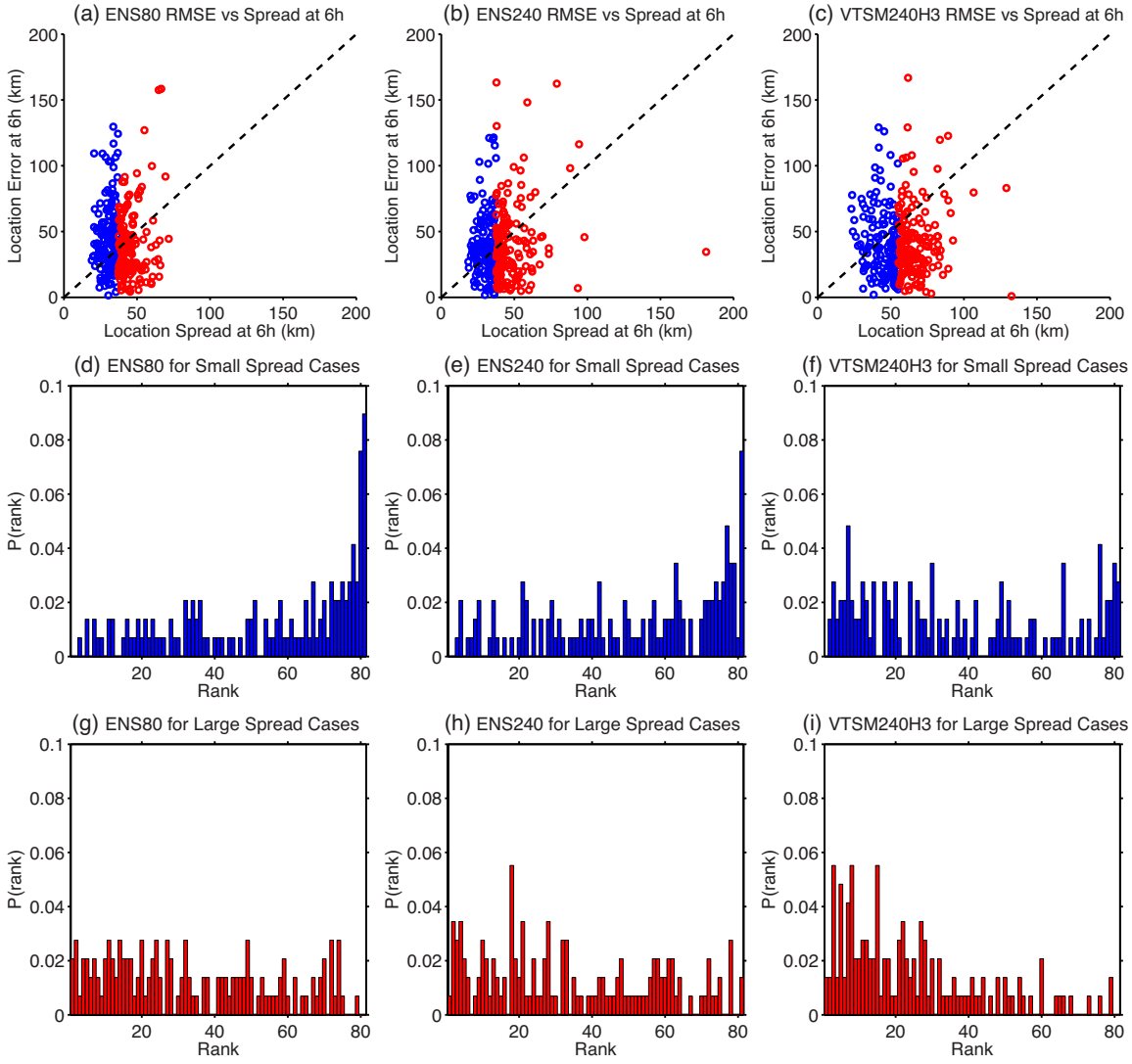

Figure 3.14 (a)-(c) Scatterplots of the 6-hour background track errors on the vertical axis against the 6-hour background ensemble track spread on the horizontal axis for (left) ENS80, (middle) ENS240 and (right) VTSM240H3. Blue (red) circles in (a)-(c) denote the equally populated samples representing small (large) background ensemble track spread in each experiment. The black dashed line is the diagonal line. (d)-(i) Rank histogram plots of (left) ENS80, (middle) ENS240 and (right) VTSM240H3 created from the samples representing (d)-(f) small background ensemble track spread and (g)-(i) large background ensemble track spread which correspond to the blue and red circles in (a)-(c), respectively, (see detailed descriptions of the rank histogram plots in the texts).

3.6 Cost comparison with ENS80

As shown in Table 3.1, compared to ENS80, the increased computational cost in VTSM240Hτ or VTSP240Hτ is only incurred by ingesting 240 members instead of 80 members during the 4DEnVar variational update and extending the 80-member background ensemble forecasts with additional τ hours. Table 3.2 shows the cost for each of the four components in a single 4DEnVar DA cycle in each experiment. The cost in each of the four components in each experiment was estimated by the wall clock time from running the same number of cores on the same type of node on the National Oceanic and Atmospheric Administration (NOAA) High Performance Computing System Jet machine. Compared to ENS80, the cost of ENS240 almost doubles in the 4DEnVar update and triples in the EnKF update and the ensemble background forecasts. In addition to the similar cost increase in 4DEnVar update as ENS240, VTSM240Hτ and VTSP240Hτ only increase the cost in the ensemble background forecasts by 9%, 14% and 20% for 1-, 2- and 3-hour shifting time interval, respectively. Overall, as shown in the last column in Table 3.2, in contrast to ENS80, the total cost in ENS240 is increased by 160%, while VTSM240Hτ and VTSP240Hτ only increase the cost by 23%, 25% and 27% for 1-, 2- and 3hour shifting time interval, respectively. Sections 3.4 and 3.5 show that VTSP240H3 improves the 6-hour temperature and wind forecasts by more than 50% and 60%, respectively, relative to the improvement in ENS240, and produces TC track forecasts with comparable or only slightly reduced skills compared to ENS240. VTSM240H3 even shows more accurate TC track forecasts than ENS240. These performance and cost results suggest that the VTS methods provide a cost-effective means to treat the sampling errors in the ensemble-based data assimilation system.

Table 3.2 Wall clock time in minutes for each of the four components in a single 4DEnVar DA cycle. The wall clock time is estimated from running on the same xJet node on the National Oceanic and Atmospheric Administration (NOAA) High Performance Computing System Jet machine. The same number of 480 cores was used in each component for different experiments

Expts	Wall clock time in minutes in each of the four components in a single 4DEnvar DA cycle				Total wall clock time in minutes	Total cost ratio relative to ENS80
	EnVar update	EnKF update	Control background Forecasts	Ensemble background forecasts		
ENS80	25	35	5	35	100	1.0
ENS240	45	110	5	100	260	2.6
VTSM240H1 or VTSP240H1	45	35	5	38	123	1.23
VTSM240H2 or VTSP240H2	45	35	5	40	125	1.25
VTSM240H3 or VTSP240H3	45	35	5	42	127	1.27

3.7 Conclusion and discussion

Instead of directly increasing the ensemble size, the VTSM and VTSP methods are implemented and explored as inexpensive means to populate the background ensemble in the NCEP GFS hybrid 4DEnVar system. The VTSM method directly takes advantage of the ensemble members at different valid times to populate the background ensemble at the analysis time. It may have the advantage of better sampling timing/phase error. By the design of shifting the ensemble perturbations at different valid times to the analysis time, the VTSP method performs temporal smoothing on the ensemble covariances, therefore eliminating the spurious covariances caused by sampling errors. To study the impacts of the VTSM and VTSP methods

in a global modeling system featured with different scales and predictabilities, both are evaluated for the general global forecasts and for the tropical cyclone forecasts in the GFS hybrid 4DEnVar system. By applying one single shifting time interval (τ=1, 2 or 3 hours), VTSM240Hτ and VTSP240Hτ triple the background ensemble size from 80 (ESN80) to 240 in the 4DEnVar variational update. ENS240 with directly running 240 members is designed as the reference to evaluate the effectiveness of the inexpensive VTSM240Hτ and VTSP240Hτ experiments.

VTSP240Hτ generally improves the global temperature and wind forecasts to 5 days. Verified against rawinsonde observations, more than 50% and 60% of the improvement from ENS240 is recovered by the best-performing VTSP experiment (VTSP240H3) for the 6-hour temperature and wind forecasts, respectively. Verified against the ECMWF analysis, VTSP240H2 produces the most consistent improvement for the temperature and wind forecasts to 5 days. Detailed diagnostics reveal that the improved global forecasts in VTSP240Hτ can be attributed to the populated background ensemble being closer to Gaussian distribution, improved accuracy of ensemble-estimated background error correlations and increased effective rank (see summary in Table 3.3). VTSP240H3 overall shows better performance than VTSP240H2 in improving the ensemble correlation accuracy and increasing the effective rank, this is consistent with the more accurate 6-hour global forecasts in VTSP240H3 (verified against both the rawinsonde observations and ECMWF analysis). On the other hand, in the global forecast verification against ECMWF analysis, the reduced forecast skills of VTSP240H3 relative to VTSP240H2 at longer lead times suggest that a tradeoff in the VTSP method needs to be taken into account. This tradeoff is between the loss of the flow-dependent features (e.g., eliminating the small-scale signals) and the gain of alleviating the sampling errors (e.g., removing the small-

scale noises), owing to the smoothing impact of the VTSP method. VTSP240H2 may achieve a better balance between these two factors. In VTSP240H3, however, the benefits of alleviating the sampling errors may dominate in the short lead times and contribute to its most improved 6-hour global forecasts, while the loss of the flow-dependency or the small-scale signals may explain its neutral impacts on the global forecasts at longer lead times.

Table 3.3 Summary of impacts of the VTSM and VTSP methods on different aspects in a global context

Impacts of VTSM and VTSP in a global context compared to ENS80		
	VTSM	VTSP
Gaussianity of background ensemble distribution	↓	↑
Ensemble correlation accuracy	↓	↑
Ensemble spread	↑	O
Effective rank	↑	↑
Global forecasts	↓ (in the troposphere)	↑

↑: improved or increased; ↓: degraded or decreased; O: neutral effect.

In contrast to VTSP240Hτ, VTSM240Hτ shows degraded global forecasts in the troposphere especially with a larger shifting time interval, VTSM240H3. This degradation may be attributed to degraded ensemble correlation accuracy, increased spread at such levels and deviation from Gaussianity in the VTSM-populated ensemble (see summary in Table 3.3). The

improved global forecasts in VTSM240Hτ above 100 hPa may be caused by the increased spread that alleviates the under-dispersiveness of the original 80-member background ensemble at such levels (see summary in Table 3.3). By comparing the components of the VTSM- and VTSP-populated background ensemble error covariances shown in Eqs. (A1) and (A3) in the appendix A, the different impacts of the VTSM and VTSP methods on those aspects in Table 3 are caused by the inclusion of the ensemble mean differences between the original and shifted ensembles in the VTSM-populated ensemble error covariances. This also suggests that the ensemble mean differences between the original and shifted ensembles fail to appropriately sample the background errors in a global forecast context.

For the TC track forecasts, experiments of ENS240, VTSM240Hτ and VTSP240Hτ are all able to improve over ENS80. The performance of VTSP240Hτ does not show strong sensitivity to the shifting time intervals. Although much less costly, VTSP240Hτ produces comparable or slightly less accurate TC track forecasts than ENS240 within the 4-day lead times and even outperforms ENS240 beyond the 4-day lead time. Like ENS240, the improved TC track forecasts in VTSP240Hτ may be originated from the improved accuracy of the estimated ensemble covariances. Larger shifting time interval in VTSM240Hτ shows enhanced improvement for the TC track forecasts. Especially, though much less costly, VTSM240H3 even shows more accurate track forecasts than ENS240. Further diagnostics suggest that the best performance of VTSM240H3 among all the experiments may be caused by its superior capability of capturing the errors from the missing sources, which is featured with the increased spread and therefore improves the reliability of the ensemble for the cases with small ensemble track spread.

Compared to ENS80, ENS240 increases the cost by 260%, while the cost in VTSM240Hτ and VTSP240Hτ is only increased by 23%, 25% and 27% for τ=1, 2, and 3 hours, respectively. Therefore, these results suggest VTSM240Hτ and VTSP240Hτ provide cost-effective ways to improve sampling errors in ensemble-based data assimilation.

As discussed in section 3.3, Lei and Whitaker (2017) found little sensitivity of forecast performance with further tuned localization scales in a similar 4DEnVar setting. Therefore, our experiments of ENS240, VTSM240Hτ and VTSP240Hτ apply the same localization length scales as in ENS80. Lorenc (2017), on the other hand, shows that increasing horizontal and vertical localization length scales are beneficial for direct increase of ensemble size and for using the time-lagged and time-shifted perturbation method to increase ensemble size in the Met Office's hybrid 4DEnVar system. The different response to the localization length scales for these two hybrid 4DEnVar systems when increasing the ensemble size is likely due to the different EnKF methods and different baseline ensemble sizes used to generate the ensembles. EnSRF with sequential assimilation and running 80 members is implemented in the GFS hybrid 4DEnVar system while the Met Office' hybrid 4DEnVar system adopts the ensemble transform Kalman filer (ETKF, Bishop et al. 2001, Wang and Bishop 2003, Wang et al. 2004 and Wang et al. 2007a) and runs 44 members.

Overall, VTSP240H2 shows the most consistent improvement for both the global forecasts and storm track forecasts in the current experiment settings. VTSM240H3 shows the best hurricane track forecasts among all the experiments whereas it generally degrades global forecasts in the troposphere. The impacts of further increasing the shifting time interval more than 3 hours in VTSM on the hurricane track forecasts remains to be investigated in the future

work by adapting the time-lagged and time-shifted perturbation method to use the form of full ensemble members in Lorenc (2017). These results also illustrate a challenge of optimizing DA algorithm in a multi-scale data assimilation system. Additional experiment is warranted when experimenting the methods in operational GFS hybrid 4DEnVar system where a wider range of scales are resolved with a higher resolution (T1534/T574).

Another set of experiments were also attempted to inexpensively increase the ensemble size within the GFS hybrid 4DEnVar system by collecting the perturbations of the ensemble forecasts that are valid at the same analysis time but initialized from previous different cycles, i.e., the time-lagged approach with using the ensemble perturbations (not shown here). Compared to the baseline experiment ENS80 as discussed in this study, this approach showed minimum or even negative impact on the global forecasts in the troposphere and the hurricane track forecasts. The only significant improvement from this approach was found for the global forecasts in the stratosphere above 100 hPa as seen in the VTSM experiments. The improved global forecasts in the stratosphere could be attributed to the increased spread by utilizing the ensemble perturbations at longer lead times which alleviates the under-dispersiveness of the original 80-member ensemble in the stratosphere as shown in Fig. 3.10. Given the inferior performance of this time-lagged approach, only the time-shifted approach is discussed in the study.

In this chapter, the paired *t*-test was applied to examine statistical significance. However, the samples do not satisfy the normal distribution and are lack of independency that are assumed in the paired *t*-test. The significance results also suffer from the multiplicity test issue (Wilks 2006). As a result, the number of current significance findings from the paired *t*-test in Figs. 3.4,

3.5, 3.6 and 3.13 in this chapter is likely reduced. Following Chapter 5, a more rigorous permutation test (Manly 2006) combined with false discovery rate method (Wilks 2006) is planned in the future to further examine the statistical significance in this chapter.

Chapter 4: The High-rank Ensemble Transform Kalman Filter

4.1 Introduction

The ensemble Kalman filter (EnKF; Evensen et al. 1994) has been widely used in the atmospheric applications, since it was introduced as a Monte Carlo realization of the traditional Kalman filter (Kalman and Bucy 1961). In the EnKF, the background error covariances are estimated and evolved by cycling an ensemble of short-range forecasts and analyses. Compared to the three-dimensional variational (3DVar) method generally employing the static background error covariances, the EnKF embraces the advantage of accounting for the flow-dependency of the forecast errors. The EnKF is therefore able to estimate the spatial, temporal and multivariate error covariances in a more realistic fashion. Different variants of the EnKF have been developed for efficient implementation purposes (Houtekamer and Mitchell 1998; Anderson 2001; Bishop et al. 2001, 2015, 2017; Whitaker and Hamill 2002; Wang and Bishop 2003; Wang et al. 2004; Hunt et al. 2007).

In the EnKF, the ensemble background error covariances, along with the observation error covariances, determine the pattern and magnitude of the corrections made on the model state variables by assimilating observations. Due to the computational constraints, the current operational EnKF systems generally run an ensemble with a size much smaller than the dimension of the numerical models (Houtekamer and Zhang 2016, Table 1). This limited ensemble size causes sampling errors and rank deficiency in the estimated background error covariance matrix. If not properly treated, these issues will incur noisy analysis increments and even filter divergence (Hamill 2006). Directly increasing the ensemble size is able to improve

the estimate of the background error covariances and thus the accuracy of the analyses and subsequent forecasts (Miyoshi et al. 2014; Lei and Whitaker 2017; Huang and Wang 2018). But the cost is very expensive. On the other hand, increasing the ensemble size in a cost-effective fashion in the ensemble-based data assimilaton system is explored and demonstrated to improve the analyses and forecasts for the storm and global scales (e.g. Xu et al. 2008; Lorenc 2017; Huang and Wang 2018).

Alternatively, the covariance localization is commonly applied in the EnKF to deal with the aforementioned issues from running a small ensemble. Its general idea is to reduce or remove the correlations between two distant variables that are assumed to be physically small or spurious. On this basis, the distance-dependent localization is applied either on the background error covariance matrix (hereafter referred to as the B-localization method) or on the observation error covariance matrix (hereafter referred to as the R-localization method). The notations of the B-localization and R-localization methods are adapted from Greybush et al. (2011) and Holland and Wang (2013). The B-localization method is typically realized through a Schur product between the raw background error covariance matrix and a predefined distance-dependent localization matrix (Houtekamer and Mitchell 2001, 2005). The R-localization method is applied through inflating the observation error variances (Hunt et al. 2007). As a result, the corrections made by the distant observations are reduced or even removed after applying the localization. In general, the distance-dependent localization function is defined to be spatially homogeneous and temporally constant. Advanced localization methods were developed in the recent studies to account for the scale, spatial or temporal dependency (Anderson 2007; Bishop and Hodyss 2007; Buehner and Charron 2007; Anderson and Lei 2013; Gasperoni and Wang 2015).

Miyoshi and Yamane (2007) and Greybush et al. (2011) found that in the assimilation of a single observation, the effective localization length scale in the R-localized Kalman gain was wider than that in the B-localized Kalman gain by applying the same localization function. It was also mentioned in these studies that the mathematical differences between the B-localization and R-localization methods were not straightforward to conclude in the assimilation of multiple observations. Sakov and Bertino (2011) compared the structures of the B-localized and R-localized Kalman gains at a single grid point, and suggested that both localization methods were expected to yield similar results in the practical applications.

While theoretical demonstrations of the mathematical differences between these two localization methods are limited, in the early studies, the performances of the B-localization and R-localization methods were usually empirically evaluated and compared in terms of the analysis accuracy by running cycled data assimilation experiments. Janjić et al. (2011) and Nerger et al. (2012) using the Lorenz-96 model (Lorenz 1996) found that the B-localization method outperformed the R-localization method especially when the observation errors were much smaller compared to the background errors. Cycled data assimilation experiments in a simplified dynamical model in Greybush et al. (2011) showed that the B-localization and R-localization methods performed comparably, if both were optimally tuned. In these studies, the B-localization method was typically applied for the variants of the serial square-root filter, and the R-localization method for the variants of the parallel implementation of the local ensemble filter. An exception was Janjić et al. (2011) which compared the B-localization and R-localization methods by performing a local analysis update using the same singular evolutive interpolated Kalman (SEIK) filter. Holland and Wang (2013) using a two-layer primitive-equation model

compared the B-localization and R-localization methods in both the same serial and same simultaneous square root filters. They found that the B-localization and R-localization methods resulted in different amounts of imbalance, which in turn affected the analysis accuracy.

This study contributes to the theoretical understanding of the differences between the B-localization and R-localization methods. A mathematical derivation is first provided with a focus on demonstrating the effective ranks of the background error covariance matrices by applying these two localization methods. The derivation does not rely on the assimilation of a single observation. Briefly, it is mathematically demonstrated in section 4.3 that for the same effective localization function, the B-localization method achieves a higher rank than the R-localization method in the localized background error covariance matrix. Meanwhile, the mathematical demonstration also shows that the B-localization method can be realized through extending and *M*odulating the raw background ensemble *P*erturbations (hereafter referred to as the MP-localization method). To reduce the computational cost, truncation of the eigenvectors from the B-localization matrix is applied to generate the modulation functions in the MP-localization method following Bishop et al. (2017). The MP-localized background error covariance matrix is thus consistent with that applying the traditional B-localization method.

The R-localization method is commonly applied in the ensemble transform Kalman filter (ETKF; Bishop et al. 2001; Wang and Bishop 2003; Wang et al. 2004) by increasing the observation error variances with an increasing distance from the model state variable (Hunt et al. 2007). In this study, the mathematical demonstration also shows that the R-localization method can be expressed in the form of the modulated background ensemble perturbations as in the B-/MP-localization method. This inspires the comparison of these two localization methods within

the same ETKF algorithm through cycled data assimilation experiments. In contrast to most of the early studies that compared these two localization methods using different filters, such a comparison of the B-localization and R-localization methods within the same ETKF algorithm makes it more straightforward to link the resulting analysis performances with the localization differences.

To emphasize the mathematically derived higher rank feature from the B-/MP-localization method, the B-/MP-localized ETKF in this study is interchangeably referred to as the high-rank ETKF (hereafter referred to as the HETKF), to distinguish it from the classic R-localized ETKF. In addition, two analysis ensemble perturbation sub-selection methods in Bishop et al. (2017) were implemented in the HETKF to investigate if such perturbation sub-selection methods affect the performances of the HETKF and R-localized ETKF.

The chapter is organized as follows. Section 4.2 briefly introduces the B-localization and R-localization methods in the context of the generic EnKF update equations. Section 4.3 provides a mathematical derivation to demonstrate the rank differences of the B-localized and R-localized Kalman gains in the generic EnKF context. The ETKF algorithm and its R-localized form are briefly described in section 4.4. Section 4.5 describes the implementation of the B-/MP-localization method in the HETKF. The performances of the HETKF and R-localized ETKF are evaluated and compared using the Lorenz model II in section 4. 6. Conclusion and discussion are presented in section 4.7.

4.2 B-localization and R-localization methods in the generic EnKF

In this section, the notations in Ide et al. (1997) are used to illustrate the generic EnKF equations (Evensen et al. 1994). The analysis $\mathbf{x}^a = [x_1^a, x_2^a, ..., x_n^a]^{\mathrm{T}}$ is obtained by correcting the background $\mathbf{x}^b = [x_1^b, x_2^b, ..., x_n^b]^{\mathrm{T}}$ with the observations $\mathbf{y}^o = [y_1^o, y_2^o, ..., y_p^o]^{\mathrm{T}}$ weighted by the Kalman gain $\mathbf{K}$,

$$\mathbf{x}^a = \mathbf{x}^b + \mathbf{K}[\mathbf{y}^o - H(\mathbf{x}^b)], \tag{4.1}$$

and

$$\mathbf{K} = \mathbf{P}^b\mathbf{H}^{\mathrm{T}}(\mathbf{H}\mathbf{P}^b\mathbf{H}^{\mathrm{T}} + \mathbf{R})^{-1}, \tag{4.2}$$

where $\mathbf{x}^a$ and $\mathbf{x}^b$ are the analysis and background vectors with a dimension of $n \times 1$, respectively; $\mathbf{y}^o$ is the observation vector with a dimension of $p \times 1$; H and $\mathbf{H}$ are the nonlinear and linearized observation operators, respectively; $\mathbf{P}^b$ denotes the background error covariance matrix with a dimension of $n \times n$; and $\mathbf{R} = \mathbf{I}r^2$ is the diagonal observation error covariance matrix with a dimension of $p \times p$. For simplicity, all the diagonal elements in $\mathbf{R}$ are set equal to r^2. The Kalman gain matrix $\mathbf{K}$ has a dimension of $n \times p$. The superscripts *a*, *b* and *o* denote the analysis, background and observations, respectively. In the EnKF, $\mathbf{P}^b$ is estimated from a *K*-member ensemble of background forecasts,

$$\mathbf{P}^b = \frac{1}{K-1}\mathbf{X}'^b(\mathbf{X}'^b)^{\mathrm{T}} = \mathbf{Z}^b(\mathbf{Z}^b)^{\mathrm{T}}, \tag{4.3}$$

where K is the ensemble size; $\mathbf{X}'^b = [\mathbf{x}_1'^b, \mathbf{x}_2'^b, ..., \mathbf{x}_K'^b] = [\mathbf{x}_1^b - \overline{\mathbf{x}}^b, \mathbf{x}_2^b - \overline{\mathbf{x}}^b, ..., \mathbf{x}_K^b - \overline{\mathbf{x}}^b]$ is the background ensemble perturbation matrix with a dimension of $n \times K$ and each column

represents the k^{th} ensemble perturbation vector $\mathbf{x}_k'^b$ with a dimension of $n\times 1$; $\overline{\mathbf{x}}^b = \sum_{k=1}^{K} \mathbf{x}_k^b \Big/ K$ is the background ensemble mean vector with a dimension of $n\times 1$; and $\mathbf{Z}^b = \mathbf{X}'^b \big/ \sqrt{K-1}$ is the background ensemble perturbation matrix normalized by a factor of $\sqrt{K-1}$.

As discussed in the introduction, the B-localization method is generally realized through a Schur product (denoted by "$\circ$") between the raw background error covariance matrix and a B-localization matrix $\mathbf{L}$ with the same dimension of $n\times n$,

$$\mathbf{P}_{\text{Bloc}}^b = \mathbf{P}^b \circ \mathbf{L}. \tag{4.4}$$

In practice, $\mathbf{L}$ is commonly defined by the Gaussian functions with its diagonal elements equal to 1.0 (see details in next section). The B-localized $\mathbf{P}_{\text{Bloc}}^b$ is only calculated once and used to update the variables at all model grid points. For example, at the i^{th} grid point, the B-localized Kalman gain calculated from $\mathbf{P}_{\text{Bloc}}^b$ is given by,

$$(\mathbf{K}_{\text{Bloc}})_i = (\mathbf{P}_{\text{Bloc}}^b \mathbf{H}^{\text{T}})_i (\mathbf{H}\mathbf{P}_{\text{Bloc}}^b \mathbf{H}^{\text{T}} + \mathbf{R})^{-1}. \tag{4.5}$$

To be consistent, in the rest of section 4, the subscript i outside the parentheses denotes the i^{th} row of a matrix or the i^{th} element of a vector. In Eq. (4.5), it denotes the i^{th} row of the matrices.

In the R-localization method, to update the variables at the i^{th} grid point, the diagonal elements in the original observation error covariance matrix $\mathbf{R}$ are inflated. A larger inflation coefficient is applied with an increasing distance away from the i^{th} grid point,

$$\mathbf{R}_{\text{Rloc}}^i = \left[\text{diag}\left(\mathbf{g}_i\right)\right]^{-1} \mathbf{R} \left[\text{diag}\left(\mathbf{g}_i\right)\right]^{-1}. \tag{4.6}$$

In Eq. (4.6), the superscript i in $\mathbf{R}^{i}_{\mathrm{Rloc}}$ denotes that the variables to be updated are at the i^{th} grid point. The vector $\mathbf{g}_i$ with a dimension of $p\times 1$ is a distance-dependent monotonically-decreasing function. It has the maximum value of 1.0 at the location of the i^{th} grid point. The vector $\mathbf{g}_i$ is commonly defined by a Gaussian function (see details in next section). "diag" is an operator that converts a vector to a diagonal matrix by aligning the elements of the vector along the diagonal. The R-localized Kalman gain at the i^{th} grid point is given by,

$$\left(\mathbf{K}_{\mathrm{Rloc}}\right)_i = \left(\mathbf{P}^b\mathbf{H}^{\mathrm{T}}\right)_i \left(\mathbf{H}\mathbf{P}^b\mathbf{H}^{\mathrm{T}} + \mathbf{R}^{i}_{\mathrm{Rloc}}\right)^{-1} \tag{4.7}$$

where the subscript i outside the parentheses, as defined earlier, denotes the i^{th} row in the matrices.

4.3 Mathematical demonstration of the higher rank of the B-localization method over the R-localization method

In this section, the B-localized and R-localized Kalman gains at the i^{th} grid point shown in section 4.2 are reformulated to examine their differences. To make the derivations in both localization methods straightforward and consistent, two assumptions are made: (i) all the model grid points are observed (i.e. n=p), and (ii) the periodic boundary condition is applied.

First, the Gaussian function $\mathbf{g}_i$ at the i^{th} grid point is defined. It determines the correlations between the i^{th} grid point and the other grid points. The following describes the formation of the B-localization matrix $\mathbf{L}$ that is used to localize the full background error covariance matrix. Since the full background error covariance matrix is associated with all the model grid points, a Gaussian matrix $\mathbf{G}=[\mathbf{g}_1,\mathbf{g}_2,\mathbf{g}_3,...,\mathbf{g}_n]$ with a dimension of $n\times n$ is first

formed, where the i^{th} column is defined by the vector $\mathbf{g}_i$ with a dimension of $n\times1$. Further following Eqs. (23) and (27) in Bishop et al. (2015), the Gaussian matrix $\mathbf{G}$ is explicitly calculated as

$$\mathbf{G}=\mathbf{F}\mathbf{\Phi}\mathbf{F}^{\mathrm{T}}, \tag{4.8}$$

where $\mathbf{F}$ is an orthonormal matrix with a dimension of $n\times n$ ($\mathbf{F}^{\mathrm{T}}\mathbf{F}=\mathbf{F}\mathbf{F}^{\mathrm{T}}=\mathbf{I}$, $\mathbf{I}$ is the identity matrix) with each column representing an eigenvector of $\mathbf{G}$, and $\mathbf{\Phi}$ is a positive semi-definite diagonal matrix with a dimension of $n\times n$ with the diagonal elements representing the eigenvalues of $\mathbf{G}$. $\mathbf{F}$ is defined by a discrete Fourier basis of sine and cosine functions following Bishop et al. (2015). The i^{th} element φ_{ii} in the diagonal matrix $\mathbf{\Phi}$ is calculated by,

$$\varphi_{ii}=n\sigma^2\frac{\exp\left\{-\left[\frac{s(i)}{d}\right]^2\right\}}{\sum_{i=1}^{n}\exp\left\{-\left[\frac{s(i)}{d}\right]^2\right\}}, \tag{4.9}$$

where $s(i)$ is the wavenumber of the i^{th} sinusoidal eigenfunction corresponding to φ_{ii} at the i^{th} grid point, and the parameter d determines the width of the distribution of the Gaussian vector $\mathbf{g}_i$. Specifically, a larger d results in a tighter Gaussian distribution. In Eq. (4.9), $n\sigma^2$ is equal to the sum of all the eigenvalues φ_{ii} in the diagonal matrix $\mathbf{\Phi}$ or the sum of all the diagonal elements in the Gaussian matrix $\mathbf{G}$. These two parameters are chosen to be $n=240$ and $\sigma^2=1$ throughout this study. Figure 1 shows an example of the distribution of the Gaussian function $\mathbf{g}_i$ defined at every 20 grid points by selecting d=3 in Eq. (4.9). The magnitude of $\mathbf{g}_i$ peaks at the i^{th} grid point (e.g., $(\mathbf{g}_i)_i=1$ where, as defined earlier, the subscript i outside the parentheses

denotes the i^{th} element in $\mathbf{g}_i$) and asymptotically decreases away from the i^{th} grid point. Here, the assumption (ii) is applied to make the Gaussian functions periodically distributed.

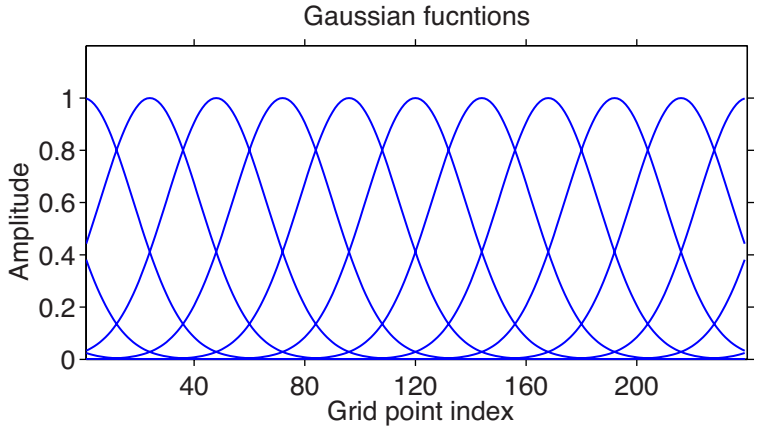

Figure 4.1 Distribution of the Gaussian functions defined at every twenty grid points.

The B-localization matrix $\mathbf{L}$ is then formed by

$$\mathbf{W} = \mathbf{G}\mathbf{G}^{\mathrm{T}}, \tag{4.10}$$

and

$$\begin{aligned}\mathbf{L} &= [\mathrm{DIAG}(\mathbf{W})]^{-1/2}\mathbf{G}\mathbf{G}^{\mathrm{T}}[\mathrm{DIAG}(\mathbf{W})]^{-1/2} \\ &= [\mathrm{DIAG}(\mathbf{W})]^{-1/2}[\mathbf{g}_1,\mathbf{g}_2,...,\mathbf{g}_n][\mathbf{g}_1,\mathbf{g}_2,...,\mathbf{g}_n]^{\mathrm{T}}[\mathrm{DIAG}(\mathbf{W})]^{-1/2}.\end{aligned} \tag{4.11}$$

where the matrix $\mathbf{W}$ has a dimension of $n \times n$ and the operator "DIAG" functions as only retaining the diagonal elements in a square matrix and setting the off-diagonal elements equal to zero. The purpose of the left- and right-multiplication of $[\mathrm{DIAG}(\mathbf{W})]^{-1/2}$ in Eq. (4.11) is to normalize the diagonal elements in the B-localization matrix $\mathbf{L}$ equal to 1.0. It can be further simplified by the periodic nature of the defined Gaussian functions,

$$\mathrm{DIAG}(\mathbf{W}) = \sum_{j=1}^{n}[\mathrm{diag}(\mathbf{g}_j)]^2 = \{\sum_{j=1}^{n}[(\mathbf{g}_j)_i]^2\}\mathbf{I} = w^2\mathbf{I}, \tag{4.12}$$

where $w = w_{ii} = \text{sqrt}\left[\sum_{j=1}^{n}[(\mathbf{g}_j)_i]^2\right]$. w_{ii} is independent of the index *i* because of the isotropic and periodic nature of the Gaussian functions.

By introducing Eqs. (4.3) and (4.11), the B-localized $\mathbf{P}^b_{\text{Bloc}}$ in Eq. (4.4) is rewritten as,

$$\begin{aligned}\mathbf{P}^b_{\text{Bloc}} &= \{\mathbf{Z}^b(\mathbf{Z}^b)^{\text{T}}\} \circ \{[\text{DIAG}(\mathbf{W})]^{-1/2}[\mathbf{g}_1,\mathbf{g}_2,...,\mathbf{g}_n][\mathbf{g}_1,\mathbf{g}_2,...,\mathbf{g}_n]^{\text{T}}[\text{DIAG}(\mathbf{W})]^{-1/2}\} \\ &= \sum_{j=1}^{n}\{[\text{DIAG}(\mathbf{W})]^{-1/2}[\text{diag}(\mathbf{g}_j)]\mathbf{Z}^b\}\{[\text{DIAG}(\mathbf{W})]^{-1/2}[\text{diag}(\mathbf{g}_j)]\mathbf{Z}^b\}^{\text{T}} \\ &= \sum_{j=1}^{n}\{[\frac{1}{w}\mathbf{I}][\text{diag}(\mathbf{g}_j)]\mathbf{Z}^b\}\{[\frac{1}{w}\mathbf{I}][\text{diag}(\mathbf{g}_j)]\mathbf{Z}^b\}^{\text{T}}.\end{aligned} \tag{4.13}$$

To simplify Eq. (4.13), we further define

$$\mathbf{V}^b_j = [\text{diag}(\mathbf{g}_j)]\mathbf{Z}^b, \tag{4.14}$$

Eq. (13) then becomes,

$$\mathbf{P}^b_{\text{Bloc}} = \frac{1}{w^2}\sum_{j=1}^{n}(\mathbf{V}^b_j)(\mathbf{V}^b_j)^{\text{T}} = \frac{1}{w^2}[\mathbf{V}^b_1,\mathbf{V}^b_2,...,\mathbf{V}^b_n][\mathbf{V}^b_1,\mathbf{V}^b_2,...,\mathbf{V}^b_n]^{\text{T}}. \tag{4.15}$$

In Eq. (4.14), the matrix $\mathbf{V}^b_j$ with a dimension of $n \times K$ can be interpreted as modulating the raw background ensemble perturbation matrix by the Gaussian function $\mathbf{g}_j$ defined at the j^{th} grid point. In particular, each column of the matrix $\mathbf{V}^b_j$ corresponds to a Schur product between a raw ensemble perturbation vector and the Gaussian vector $\mathbf{g}_j$ defined at the j^{th} grid point. Equations (4.13)-(4.15) suggest that the B-localization method can be realized by an outer-product of the expanded modulated ensemble perturbation matrix $\frac{1}{w}[\mathbf{V}^b_1,\mathbf{V}^b_2,...,\mathbf{V}^b_n]$ with a dimension of $n \times (nK)$. Since the B-localization method here is achieved through expanding and *M*odulating

the raw ensemble *P*erturbation matrix, it is interchangeably termed as the MP-localization method.

Correspondingly, the Kalman gain at the i^{th} grid point with the B-localized $\mathbf{P}^{b}_{\text{Bloc}}$ in Eq. (4.5) is reformulated by introducing Eq. (4.15),

$$\begin{aligned}(\mathbf{K}_{\text{Bloc}})_i &= \{[\frac{1}{w^2}\sum_{j=1}^{n}(\mathbf{V}_j^b)(\mathbf{V}_j^b)^{\mathrm{T}}]\mathbf{H}^{\mathrm{T}}\}_i\{\mathbf{H}[\frac{1}{w^2}\sum_{j=1}^{n}(\mathbf{V}_j^b)(\mathbf{V}_j^b)^{\mathrm{T}}]\mathbf{H}^{\mathrm{T}}+\mathbf{R}\}^{-1} \\ &= \{\frac{1}{w^2}\sum_{j=1}^{n}(\mathbf{V}_j^b)_i(\mathbf{H}\mathbf{V}_j^b)^{\mathrm{T}}\}\{\frac{1}{w^2}\sum_{j=1}^{n}(\mathbf{H}\mathbf{V}_j^b)(\mathbf{H}\mathbf{V}_j^b)^{\mathrm{T}}+\mathbf{R}\}^{-1}.\end{aligned} \tag{4.16}$$

For a more straightforward comparison with the R-localized Kalman gain form as will be introduced soon, Eq. (4.16) is further right-multiplied by an identity matrix with a dimension of $n\times n$ that is expressed as $\mathbf{I}=\sqrt{\sum_{j=1}^{n}[\text{diag}(\mathbf{g}_j)]^2}\Big/\sqrt{\sum_{j=1}^{n}[(\mathbf{g}_j)_i]^2}$ by referring to Eq. (4.12),

$$(\mathbf{K}_{\text{Bloc}})_i = \{\frac{1}{\sum_{j=1}^{n}[(\mathbf{g}_j)_i]^2}\sum_{j=1}^{n}(\mathbf{V}_j^b)_i(\mathbf{H}\mathbf{V}_j^b)^{\mathrm{T}}\}\{\frac{1}{\sum_{j=1}^{n}[(\mathbf{g}_j)_i]^2}\sum_{j=1}^{n}(\mathbf{H}\mathbf{V}_j^b)(\mathbf{H}\mathbf{V}_j^b)^{\mathrm{T}}+\mathbf{R}\}^{-1}\frac{\sqrt{\sum_{j=1}^{n}[\text{diag}(\mathbf{g}_j)]^2}}{\sqrt{\sum_{j=1}^{n}[(\mathbf{g}_j)_i]^2}}. \tag{4.17}$$

In the R-localization method, given the assumption (i) that all the model grid points are observed (i.e. *n*=*p*), the Gaussian function $\mathbf{g}_i$ at the i^{th} grid point is applied in Eq. (4.6) to calculate the localized $\mathbf{R}^{i}_{\text{Rloc}}$. The R-localized Kalman gain at the i^{th} grid point is then obtained by introducing Eqs. (4.3) and (4.6) to Eq. (4.7),

$$\begin{aligned}(\mathbf{K}_{\text{Rloc}})_i &= \{\mathbf{Z}^b(\mathbf{H}\mathbf{Z}^b)^{\mathrm{T}}\}_i\{\mathbf{H}\mathbf{Z}^b(\mathbf{H}\mathbf{Z}^b)^{\mathrm{T}}+[\text{diag}(\mathbf{g}_i)]^{-1}\mathbf{R}[\text{diag}(\mathbf{g}_i)]^{-1}\}^{-1} \\ &= (\mathbf{Z}^b)_i(\mathbf{H}[\text{diag}(\mathbf{g}_i)\mathbf{Z}^b])^{\mathrm{T}}[(\mathbf{H}[\text{diag}(\mathbf{g}_i)\mathbf{Z}^b])(\mathbf{H}[\text{diag}(\mathbf{g}_i)\mathbf{Z}^b])^{\mathrm{T}}+\mathbf{R}]^{-1}\text{diag}(\mathbf{g}_i).\end{aligned} \tag{4.18}$$

By introducing the modulated background ensemble perturbation matrix in Eq. (14), Eq. (18) is simplified as,

$$(\mathbf{K}_{\mathrm{Rloc}})_i = (\mathbf{V}_i^b)_i(\mathbf{HV}_i^b)^{\mathrm{T}}[(\mathbf{HV}_i^b)(\mathbf{HV}_i^b)^{\mathrm{T}} + \mathbf{R}]^{-1}\mathrm{diag}(\mathbf{g}_i), \tag{4.19}$$

where $(\mathbf{Z}^b)_i = (\mathbf{V}_i^b)_i$, given $(\mathbf{g}_i)_i = 1$.

To facilitate the comparison with the B-localized Kalman gain form in Eq. (4.17), Eq. (4.19) is further reformulated by introducing $[(\mathbf{g}_i)_i]^2 = 1$, $\mathrm{diag}(\mathbf{g}_i) = \sqrt{[\mathrm{diag}(\mathbf{g}_i)]^2} \Big/ \sqrt{[(\mathbf{g}_i)_i]^2}$ and the Kronecker delta function δ_{ij} in Eq. (4.21) below,

$$\begin{aligned}(\mathbf{K}_{\mathrm{Rloc}})_i &= \frac{1}{[(\mathbf{g}_i)_i]^2}(\mathbf{V}_i^b)_i(\mathbf{HV}_i^b)^{\mathrm{T}}\,[\frac{1}{[(\mathbf{g}_i)_i]^2}(\mathbf{HV}_i^b)(\mathbf{HV}_i^b)^{\mathrm{T}} + \mathbf{R}]^{-1}\frac{\sqrt{[\mathrm{diag}(\mathbf{g}_i)]^2}}{\sqrt{[(\mathbf{g}_i)_i]^2}} \\ &= \{\frac{1}{\sum_{j=1}^{n}\delta_{ij}[(\mathbf{g}_j)_i]^2}\sum_{j=1}^{n}\delta_{ij}(\mathbf{V}_j^b)_i(\mathbf{HV}_j^b)^{\mathrm{T}}\}\{\frac{1}{\sum_{j=1}^{n}\delta_{ij}[(\mathbf{g}_j)_i]^2}\sum_{j=1}^{n}\delta_{ij}(\mathbf{HV}_j^b)(\mathbf{HV}_j^b)^{\mathrm{T}} + \mathbf{R}\}^{-1}\frac{\sqrt{\sum_{j=1}^{n}\{\delta_{ij}[\mathrm{diag}(\mathbf{g}_j)]^2\}}}{\sqrt{\sum_{j=1}^{n}\delta_{ij}[(\mathbf{g}_j)_i]^2}},\end{aligned} \tag{4.20}$$

where,

$$\delta_{ij} = \begin{cases} 1, & \mathrm{i} = \mathrm{j} \\ 0, & \mathrm{i} \neq j \end{cases}. \tag{4.21}$$

Equations (4.18)-(4.20) suggest that at the i^{th} grid point, the localization effect by inflating the observation error variances in the R-localization method can be equivalently achieved by modulating the raw ensemble perturbation matrix with the Gaussian function $\mathbf{g}_i$ defined in Eq. (4.14). Such a reformulation assists in a direct mathematical comparison between the B-localization and R-localization methods.

By comparing Eqs. (4.17) and (4.20), it can be seen that the R-localization method can be regarded as a special case of the B-localization method when expressed using the Kronecker delta function in Eq. (4.21). The number of terms in the summations over the modulated ensemble perturbation matrix index j in the B-localization method in Eq. (4.17) is reduced to one in the R-localization method in Eq. (4.20). Specifically, in the B-localization method, a total of n modulated background ensemble perturbation matrices are involved in the calculation of the B-localized Kalman gain at the i^{th} grid point. However, the R-localization method only includes the contribution from a single modulated ensemble perturbation matrix associated with the Gaussian function defined at the i^{th} grid point. As a result, the rank of the B-localized Kalman gain is higher than that of the R-localization method. The above conclusion can also be drawn from a simple linear algebra analysis. Given $n>p>K$ in general, the rank of the original Kalman gain in Eq. (4.2) is K-1. It is determined by the minimum of the ranks of $\mathbf{H}$, $\mathbf{P}^b$ and $\mathbf{R}$ in Eq. (4.2) that are p, K-1 and p, respectively. Inflating the observation error variances of the original $\mathbf{R}$ as in Eq. (4.6) will not change its rank. The R-localized Kalman gain in Eq. (4.7) is thus of the same rank as the original Kalman gain. However, after applying the B-localization method, the maximum rank of $\mathbf{P}^b_{\text{Bloc}}$ in Eq. (4.4) can reach n due to its resultant blocked diagonal structure. The B-localized Kalman gain in Eq. (4.5) thus has an increased rank of p. Therefore, the linear algebra analysis also suggests a higher rank of the B-localized Kalman gain in contrast to the original and R-localized Kalman gains. This is consistent with the mathematical demonstration in this section.

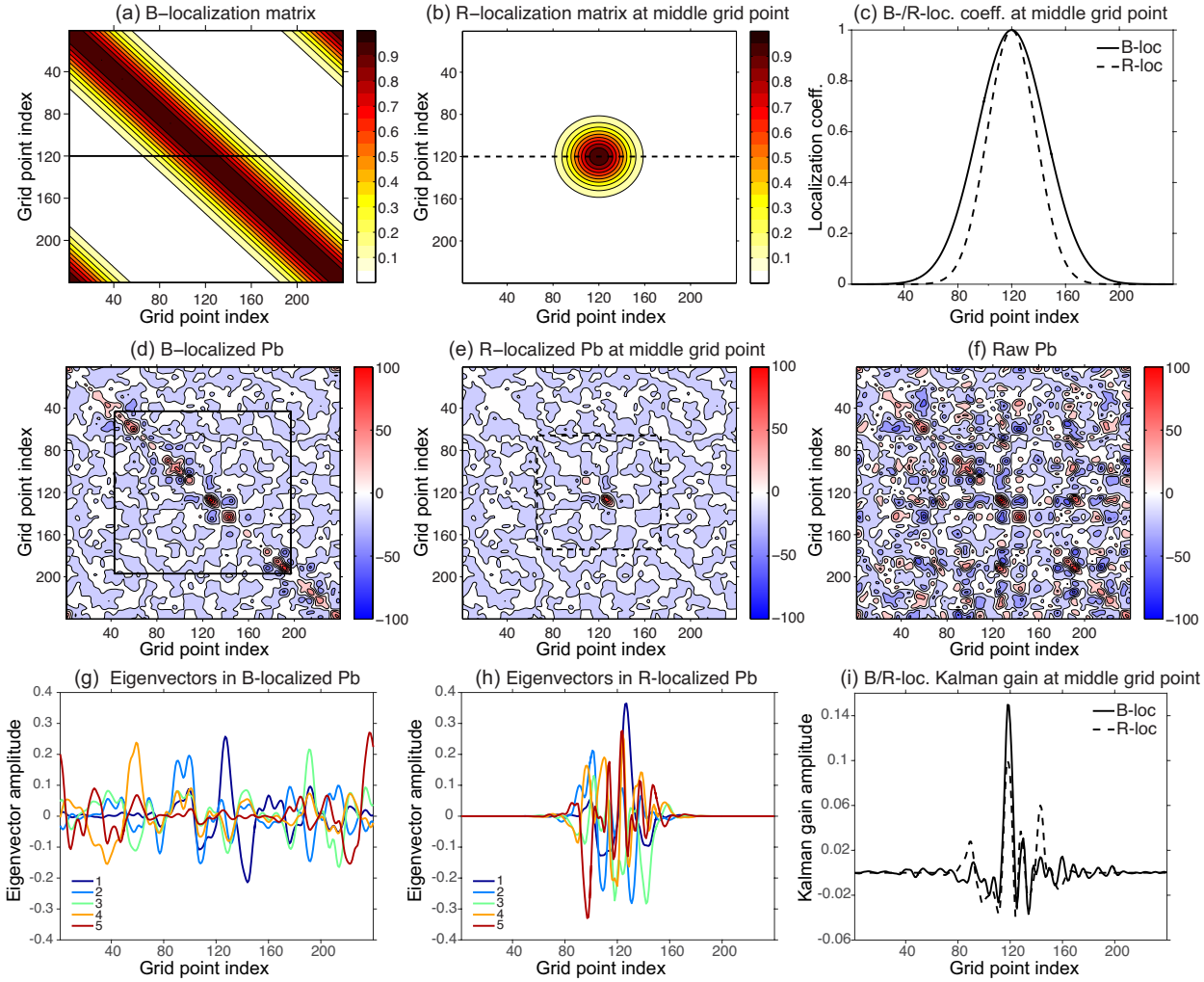

Figure 4.2 Illustration of (a) the matrix $\frac{1}{w^2}\sum_{j=1}^{n}\mathbf{g}_j\mathbf{g}_j^{\mathrm{T}}$ for the B-localization method in Eq. (4.22), (b) the matrix $\mathbf{g}_i\mathbf{g}_i^{\mathrm{T}}$ (i=120) for the R-localization method in Eq. (4.23) calculated with d=3 and n=240 in Eqs. (4.8) and (4.9), and (c) the localization coefficients from the 120$^{\text{th}}$ row of the localization matrix for the B-localization method (solid line) in (a) and the R-localization method (dashed line) in (b), respectively. The effective localization distance is defined as half of the interval where the correlation coefficients taper to 0.01. Illustration of the observation-space ensemble background error covariance matrix estimated from 6 members with (f) no localization, (d) the B-localization method and (e) the R-localization method, and the size of the black solid and dashed squares in (d) and (e), respectively, doubles the effective localization distances as shown in (c). First five leading eigenvectors (colored solid lines) decomposed from (g) the B-localized and (h) the R-localized observation-space ensemble background error covariance matrices in (d) and (e), respectively. (i) Kalman gains of updating the 120$^{\text{th}}$ grid point from the B-localization method (solid line) and the R-localization method (dashed line).

The localization effects of the B-localization and R-localization methods are further isolated and compared by reformulating their localized observation-space background error covariance matrices in Eqs. (4.17) and (4.20). For the B-localization method in Eq. (4.17),

$$\frac{1}{\sum_{j=1}^{n}[(\mathbf{g}_j)_i]^2}\sum_{j=1}^{n}(\mathbf{H}\mathbf{V}_j^b)(\mathbf{H}\mathbf{V}_j^b)^{\mathrm{T}} = \frac{1}{w^2}\sum_{j=1}^{n}\{\mathbf{H}[\mathrm{diag}(\mathbf{g}_j)\mathbf{Z}^b]\}\{\mathbf{H}[\mathrm{diag}(\mathbf{g}_j)\mathbf{Z}^b]\}^{\mathrm{T}} \\ = \{[\mathbf{H}\mathbf{Z}^b(\mathbf{H}\mathbf{Z}^b)^{\mathrm{T}}]\circ[\frac{1}{w^2}\sum_{j=1}^{n}(\mathbf{g}_j\mathbf{g}_j^{\mathrm{T}})]\}. \tag{4.22}$$

For the R-localization method in Eq. (4.20),

$$\frac{1}{\sum_{j=1}^{n}\delta_{ij}[(\mathbf{g}_j)_i]^2}\sum_{j=1}^{n}\delta_{ij}(\mathbf{H}\mathbf{V}_j^b)(\mathbf{H}\mathbf{V}_j^b)^{T} = \{\mathbf{H}[\mathrm{diag}(\mathbf{g}_i)\mathbf{Z}^b]\}(\mathbf{H}[\mathrm{diag}(\mathbf{g}_i)\mathbf{Z}^b]\}^{\mathrm{T}} \\ = \{[\mathbf{H}\mathbf{Z}^b(\mathbf{H}\mathbf{Z}^b)^{\mathrm{T}}]\circ(\mathbf{g}_i\mathbf{g}_i^{\mathrm{T}})\}. \tag{4.23}$$

Equations (4.22) and (4.23) suggest that the effects of the localization applied on the observation-space background error covariance matrices are determined by the matrix $\frac{1}{w^2}\sum_{j=1}^{n}\mathbf{g}_j\mathbf{g}_j^{\mathrm{T}}$ for the B-localization method and by the matrix $\mathbf{g}_i\mathbf{g}_i^{\mathrm{T}}$ for the R-localization method, respectively. Figure 2 shows an example of the structures of these two matrices calculated at the 120$^{\mathrm{th}}$ grid point (e.g., i=120). The Gaussian matrix $\mathbf{G}$ in Eq. (8) is calculated by selecting d=3 and n=240 as in Fig. 4.1. The matrix $\frac{1}{w^2}\sum_{j=1}^{n}\mathbf{g}_j\mathbf{g}_j^{\mathrm{T}}$ in Eq. (4.22) for the B-localization method (Fig. 4.2a) shows a symmetric structure with the magnitude equal to 1.0 on the diagonal and monotonically decreasing away from the diagonal. In contrast, the matrix $\mathbf{g}_i\mathbf{g}_i^{\mathrm{T}}$ in Eq. (4.23) for the R-localization method (Fig. 4.2b) shows a localized circular structure centered at the element

of (120, 120). For illustration purposes, the 120th column of these two matrices (e.g., the solid black line and the dashed black line in Figs. 4.2a,b) is used to determine the effective localization distance for the B-localization and R-localization methods in this study. Here the effective localization distance is defined as half of the interval where the localization coefficients taper to 0.01. As can be seen in Fig. 4.2c, the same localization parameter d=3 in Eq. (4.9) results in a broader effective localization distance in the B-localization method than that in the R-localization method. This result seems inconsistent with the expectation that the tighter effective localization distance generally results in a higher rank of the localized background error covariance matrix. However, Eqs. (4.17) and (4.20) suggest that the mathematically derived higher rank from the B-localization method is independent of the effective localization distance. Meanwhile, the effective localization distances in these two localization methods in Fig. 4.2c are caused by and consistent with the constructions of their localization matrices. Therefore, cautions need to be taken to relate the effective localization distance with the rank of the localized background error covariance matrix especially when different forms of localization are utilized.

To further verify the mathematical demonstration, the effective ranks resulted from both localization methods are calculated and compared in an example using $\mathbf{H}=\mathbf{I}$. In Fig. 4.2f, the raw observation-space background error covariance matrix is estimated from six members. Through a Schur product with the B-localization matrix (Fig. 4.2a), the B-localized observation-space background error covariance matrix retains the covariances near the diagonal and reduces or even removes the covariances away from the diagonal (Fig. 4.2d). On the other hand, due to the "local" nature of the R-localization matrix (Fig. 4.2b), the R-localized observation-space background error covariance matrix (Fig. 4.2e) have more zero values on the diagonal in contrast

to that applying the B-localization method. The E dimension (Patil et al. 2001; Oczkowski et al. 2005; Kuhl et al. 2007; Huang and Wang 2018) is used to evaluate the effective rank of the background error covariance matrix. The E dimension was calculated within a square as shown in Figs. 4.2d,e. The size of the square is defined as twice of the effective localization distance. The B-localization method increases the E dimension from 4.21 to 10.33. This is consistent with the expansion of the dimension of the modulated ensemble perturbation matrix after applying the B-localization method in Eq. (4.15). In contrast, the R-localization method results in a reduced E dimension of 2.81 from 3.72. This is likely due to the occurrence of more zero values on the diagonal of the R-localized background error covariance matrix. Therefore, the mathematically derived higher rank in the B-localized observation-space background error covariance matrix is consistently demonstrated by comparing their resulting effective ranks.

The different structures of the B-localized and R-localized observation-space background error covariance matrices further motivate to investigate how many observations would literally influence their resulting analyses. The Kalman gains at the 120^{th} grid point are thus calculated for these two localization methods. In general, the matrix inversion in Eq. (4.2) for the Kalman gain calculation can be solved by using the eigenvectors decomposed from the observation-space background error covariance matrix (Bishop et al. 2017). Figures 4.2g,h show the first five leading eigenvectors decomposed from the B-localized and R-localized observation-space background error covariance matrices in Figs. 4.2d,e, respectively. The eigenvectors from the B-localized observation-space background error covariance matrix cover the whole domain (Fig. 4.2g). However, all the five leading eigenvectors from the R-localized observation-space background error covariance matrix are confined in a local area centered at the 120^{th} grid point.

Their Kalman gains are then calculated using $\mathbf{R} = \mathbf{I}$ as shown in Fig. 4.2i. The B-localized Kalman gain at the 120th grid point extends over the whole domain with larger values at the grid points close to the 120th grid point. However, the R-localized Kalman gain is confined in a local area between the 60th and 180th grid points. This local nature of the R-localized Kalman gain is also shown in the curves representing the R-localization method in Figs. 4.2c,h. As a result, all the observations would contribute to updating a single grid point in the B-localization method, although the distant observations make less contribution. This also suggests that the optimal effective localization distance defined in this study does not explicitly suggest the number of the observations that influence the analysis in the B-localization method. In contrast, the analysis at a particular grid point in the R-localization method is influenced by limited observations that are close.

4.4 Implementation of the R-localization method in the ETKF

The mathematical demonstration in section 4.3 suggests that the traditional B-localization method in Eq. (4.4) can be realized by expanding and modulating the background ensemble perturbations through Eq. (4.15). This allows the implementation of the B-/MP-localization method in the ETKF that generally employs the R-localization method. This section first briefly describes the classic R-localized ETKF. The HETKF applying the MP-localization method will be discussed in section 4.5.

In the ensemble transform Kalman filter (ETKF; Bishop et al. 2001; Wang and Bishop 2003; Wang et al. 2004) and its local form (LETKF; Hunt et al. 2007), the background ensemble

perturbations are transformed to the analysis ensemble perturbations on the ensemble perturbation subspace by a transform matrix $\mathbf{T}$,

$$\mathbf{Z}^a = \mathbf{Z}^b\mathbf{T}, \tag{4.24}$$

where $\mathbf{Z}^a = \mathbf{X}'^a/\sqrt{K-1}$ is the analysis ensemble perturbation matrix $\mathbf{X}'^a$ with a dimension of $n \times K$ normalized by a factor of $\sqrt{K-1}$. As discussed in Bishop et al. (2001), Wang and Bishop (2003) and Wang et al. (2004), the transform matrix $\mathbf{T}$ is calculated to make sure that the analysis error covariances $\mathbf{P}^a = \mathbf{Z}^a(\mathbf{Z}^a)^{\mathrm{T}}$ are updated by satisfying the optimal data assimilation theory $\mathbf{P}^a = (\mathbf{I}-\mathbf{KH})\mathbf{P}^b$,

$$\mathbf{T} = \mathbf{C}(\mathbf{\Gamma}+\mathbf{I})^{-1/2}\mathbf{C}^{\mathrm{T}}, \tag{4.25}$$

and

$$\mathbf{A} = (\mathbf{HZ}^b)^{\mathrm{T}}\mathbf{R}^{-1}(\mathbf{HZ}^b) = \mathbf{C}\mathbf{\Gamma}\mathbf{C}^{\mathrm{T}}, \tag{4.26}$$

where each column of the matrix $\mathbf{C}$ represents an eigenvector of the matrix $\mathbf{A}$ and the diagonal matrix $\mathbf{\Gamma}$ contains the corresponding eigenvalues.

The background ensemble mean is updated by,

$$\overline{\mathbf{x}}^a = \overline{\mathbf{x}}^b + \mathbf{Z}^b[\mathbf{C}(\mathbf{\Gamma}+\mathbf{I})^{-1}\mathbf{C}^{\mathrm{T}}](\mathbf{HZ}^b)^{\mathrm{T}}\mathbf{R}^{-1}[\mathbf{y}^o - \overline{H(\mathbf{x}^b)}]. \tag{4.27}$$

where the overbar denotes the ensemble mean.

The R-localization method in the ETKF is realized by following its implementation in the LETKF of Hunt et al. (2007). The update of the model state variables is performed independently at different model grid points. At the i^{th} grid point, instead of using the original $\mathbf{R}$, the localized observation error covariance matrix $\mathbf{R}^i_{\text{Rloc}}$ defined in Eq. (6) is applied in Eqs.

(4.24)-(4.27). Different from the LETKF that selects a subset of observations, the whole set of observations are selected here to update the variables at the i^{th} grid point in the R-localized ETKF. This "global" analysis update is designed to assure a homogeneous comparison with the HETKF detailed in section 4.5.

4.5 Implementation of the high-rank ETKF (HETKF) by applying the MP-localization method

As shown in Eq. (4.15), the B-localized $\mathbf{P}^{b}_{\text{Bloc}}$ can be achieved by an outer-product of an expanded modulated background ensemble perturbation matrix $\frac{1}{w}[\mathbf{V}^{b}_{1},\mathbf{V}^{b}_{2},...,\mathbf{V}^{b}_{n}]$ with a dimension of $n\times(nK)$. This expression makes it possible to implement the B-localization method in the ETKF. However, the computational cost is very expensive, because it requires an eigen-decomposition of a matrix with a dimension of $(nK)\times(nK)$ in Eq. (4.26). To reduce the computational cost, following Bishop et al. (2017), the method of selecting the leading eigenvalues and eigenvectors of the original B-localization matrix is implemented to reduce the number of the modulation functions and thus the size of the extended modulated background ensemble.

4.5.1 Specific implementation of the MP-localization method in the ETKF (HETKF)

Instead of directly using the columns of the Gaussian matrix $\mathbf{G}$ that forms the B-localization matrix $\mathbf{L}$ as the modulation functions in Eq. (4.14), the modulation functions in Bishop et al. (2017) are calculated from the leading eigenvalues and eigenvectors of the B-

localization matrix $\mathbf{L}$. As a result, the size of the modulated background ensemble perturbation matrix is significantly reduced compared to the matrix $\frac{1}{w}[\mathbf{V}_1^b, \mathbf{V}_2^b, ..., \mathbf{V}_n^b]$ in Eq. (4.15). The procedures are detailed as follows.

(i) Calculate the eigenvalues and eigenvectors of the original B-localization matrix $\mathbf{L}$ and order them correspondingly from the largest to the smallest eigenvalue. In Eq. (4.28), the diagonal matrix $\mathbf{\Lambda}$ contains the eigenvalues of the B-localization matrix $\mathbf{L}$ in a descending order and the columns of the matrix $\mathbf{E}$ represent the corresponding eigenvectors.

$$\mathbf{L} = \mathbf{E}\mathbf{\Lambda}\mathbf{E}^{\mathrm{T}} = (\mathbf{E}\mathbf{\Lambda}^{1/2})(\mathbf{E}\mathbf{\Lambda}^{1/2})^{\mathrm{T}}. \tag{4.28}$$

(ii) Calculate the modulation matrix $\hat{\mathbf{G}}$ by selecting and normalizing the first M leading eigenvalues and eigenvectors to form the localization matrix $\mathbf{L}_{\mathrm{MP}}$. In this study, the first M leading eigenvalues and eigenvectors are selected to account for more than 99% of the sum of all the eigenvalues following Bishop et al. (2017). Mathematically,

$$\mathbf{L}_{1\sim M} = (\mathbf{E}\mathbf{\Lambda}^{1/2})_{1\sim M}(\mathbf{E}\mathbf{\Lambda}^{1/2})_{1\sim M}^{\mathrm{T}}, \tag{4.29}$$

$$\hat{\mathbf{G}} = \{[\mathrm{diag}(\mathbf{L}_{1\sim M})]^{-1/2}\left[(\mathbf{E}\mathbf{\Lambda}^{1/2})\right]_{1\sim M}\} = [\hat{\mathbf{g}}_1, \hat{\mathbf{g}}_2, ..., \hat{\mathbf{g}}_M], \tag{4.30}$$

and

$$\mathbf{L}_{\mathrm{MP}} = \hat{\mathbf{G}}\hat{\mathbf{G}}^{\mathrm{T}}. \tag{4.31}$$

(iii) Generate an expanded modulated background ensemble perturbation matrix $\hat{\mathbf{X}}'^b$ with a dimension of $n \times (MK)$ by a Schur product between each raw ensemble perturbation vector and each column of the modulation matrix $\hat{\mathbf{G}}$. Mathematically,

$$\hat{\mathbf{X}}'^{b} = \sqrt{\frac{MK-1}{K-1}}[\mathrm{diag}(\hat{\mathbf{g}}_1)\mathbf{X}'^{b}, \mathrm{diag}(\hat{\mathbf{g}}_2)\mathbf{X}'^{b}, ..., \mathrm{diag}(\hat{\mathbf{g}}_M)\mathbf{X}'^{b}], \tag{4.32}$$

and

$$\hat{\mathbf{Z}}^{b} = \frac{\hat{\mathbf{X}}'^{b}}{\sqrt{MK-1}}. \tag{4.33}$$

The expanded modulated background ensemble can be obtained by adding the modulated ensemble perturbations in Eq. (4.32) to the original background ensemble mean. As such, the modulated background ensemble shares the same background ensemble mean as the original ensemble. The factor $\sqrt{MK-1}/\sqrt{K-1}$ in Eq. (4.32) is to ensure that the localized ensemble background error covariance matrices with and without the modulated perturbation form are equivalent, e.g.,

$$\frac{\hat{\mathbf{X}}'^{b}(\hat{\mathbf{X}}'^{b})^{\mathrm{T}}}{MK-1} = \frac{\mathbf{X}'^{b}(\mathbf{X}'^{b})^{\mathrm{T}}}{K-1} \circ \mathbf{L}_{\mathrm{MP}} \cong \frac{\mathbf{X}'^{b}(\mathbf{X}'^{b})^{\mathrm{T}}}{K-1} \circ \mathbf{L}. \tag{4.34}$$

For the computational concern, only M modulation functions are selected and used in the implementation of the MP-localization method, in contrast to using n modulation functions as in Eq. (4.15). In general, M is expected to be much smaller than n. But the M modulation functions are constructed to account for more than 99% of the variances of the original B-localization matrix. As suggested in Bishop et al. (2017), it is expected to cause minimum effects on the resulting effective rank of the localized background error covariance matrix and the cycled data assimilation experiment results in section 4.6. More importantly, the use of fewer modulation functions in the MP-localization method can significantly improve the computational efficiency.

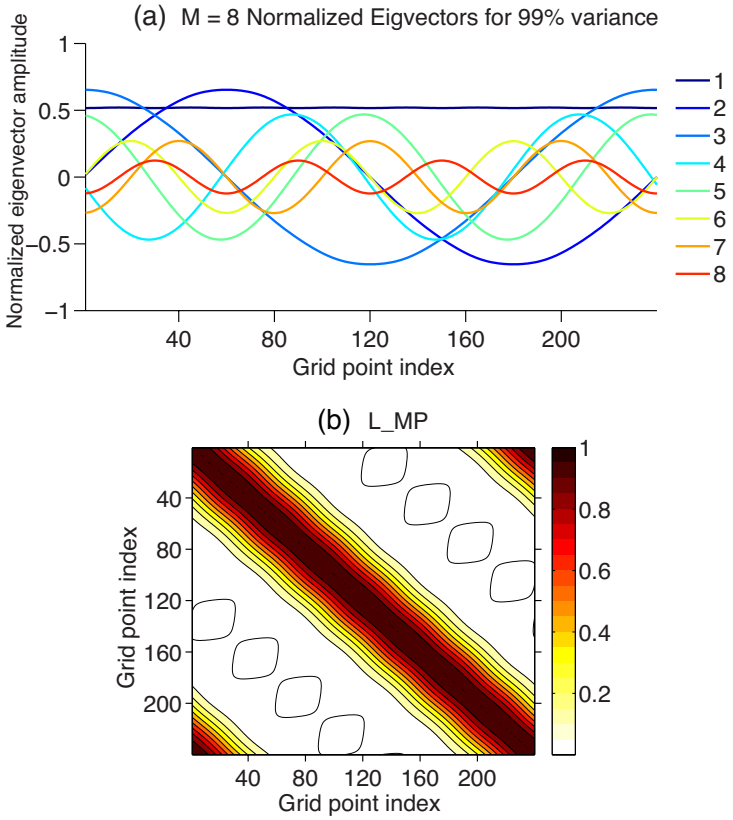

Figure 4.3 Illustration of (a) the eight modulation functions (colored solid lines) in $\hat{\mathbf{G}}$ and (b) the associated localization matrix $\mathbf{L}_{\mathrm{MP}}$ calculated from the B-localization matrix in Fig. 4.2a.

Figure 4.3 shows an example of the modulation functions in the matrix $\hat{\mathbf{G}}$ and the associated localization matrix $\mathbf{L}_{\mathrm{MP}}$ calculated from the original B-localization matrix $\mathbf{L}$ in Fig. 4.2a. In this example, the first eight leading eigenvalues and corresponding eigenvectors are selected to account for more than 99% of the sum of all the eigenvalues of the original B-localization matrix $\mathbf{L}$. The resulting localization matrix $\mathbf{L}_{\mathrm{MP}}$ (Fig. 4.3b) almost recovers the original B-localization matrix $\mathbf{L}$ (Fig. 4.2a). In particular, the size of the modulated background ensemble perturbation matrix is only increased by a factor of 8 using Eq. (4.32) instead of a factor of 240 using Eq. (4.15) in this example. This significantly improves the computational

efficiency in the HETKF. For an isotropic B-localization matrix as in Fig. 4.2a, the associated modulation function used in the MP-localization method has a larger wave number with an increasing eigenvalue (Fig. 4.3a).

4.5.2 Ensemble mean and perturbation update in the HETKF

Following section 4.4, Eqs. (4.24)-(4.27) are used for the HETKF ensemble mean and perturbation update. Instead of using $\mathbf{Z}^b$ with a dimension of $n \times K$, $\hat{\mathbf{Z}}^b$ with a dimension of $n \times (MK)$ is applied in these equations.

During the ensemble perturbation update, directly applying $\hat{\mathbf{Z}}^b$ with a dimension of $n \times (MK)$ in Eq. (4.24) would produce *MK* analysis perturbations in the HETKF. In the practical applications, the *K* analysis perturbations need to be selected to initialize a *K*-member ensemble of background forecasts before advancing to the next DA cycle. To have a robust comparison of the B-/MP-localization and R-localization methods, following Bishop et al. (2017), two methods of sub-selecting the analysis perturbations during the ensemble perturbation update in the HETKF were implemented and examined.

The first perturbation sub-selection method is defined as deterministic to distinguish it from the second, stochastic method. In the deterministic perturbation sub-selection method, the first *K* columns of $\hat{\mathbf{Z}}^a$ are selected. To remove the modulation effect, a demodulation procedure in Eq. (4.35) is applied by left-multiplying each column with a diagonal matrix $[\mathrm{diag}(\hat{\mathbf{g}}_1)]^{-1}$ associated with the first modulation function,

$$\hat{\mathbf{Z}}^a_{\mathrm{MP-D}} = \{[\mathrm{diag}(\hat{\mathbf{g}}_1)]^{-1}\hat{\mathbf{z}}^a_1, [\mathrm{diag}(\hat{\mathbf{g}}_1)]^{-1}\hat{\mathbf{z}}^a_2, ..., [\mathrm{diag}(\hat{\mathbf{g}}_1)]^{-1}\hat{\mathbf{z}}^a_K\}. \tag{4.35}$$

where the letter "D" in the subscript "MP-D" stands for "deterministic". Finally, the K-member analysis ensemble perturbation matrix $\mathbf{X}'^{a}_{\mathrm{MP-D}}$ is recovered by referring to the relation between $\hat{\mathbf{X}}'^{b}$ and $\hat{\mathbf{Z}}^{b}$ in Eqs. (4.32) and (4.33),

$$\mathbf{X}'^{a}_{\mathrm{MP-D}} = \sqrt{K-1}\hat{\mathbf{Z}}^{a}_{\mathrm{MP-D}}. \tag{4.36}$$

In Bishop et al. (2017), this deterministic perturbation sub-selection method was compared with a more robustly derived selection approach termed as the gain-form ETKF. Our further examination using the Lorenz model II (see next section for details) showed that for the isotropic B-localization matrix $\mathbf{L}$, the performances of these two perturbation selection methods were statistically indistinguishable (not shown here). So in this study we use this deterministic perturbation sub-selection method described in Eq. (4.35) for the cycled data assimilation experiments. A similar sub-selection procedure was adopted in Kretschmer et al. (2015) to select the K analysis ensemble perturbations updated from a background ensemble formed by a mixture of the flow-dependent and climatological perturbations. This HETKF implementation and the classic R-localized ETKF described in section 4.4 are hereafter referred to as "MP-D" and "R-D".

The second perturbation sub-selection method, defined as stochastic, is based on the idea of updating each member with different sets of perturbed observations (Houtekamer and Mitchell 1998). Specifically, the K sets of perturbed observations are generated and assimilated to update the raw ensemble members in the HETKF. This HETKF implementation is denoted as "MP-S" in this study, where the letter "S" stands for stochastic. To have a homogeneous comparison of the B-/MP-localization and R-localization methods, the same perturbed observation approach is also applied for the R-localized ETKF, which is denoted as "R-S" hereafter. This stochastic approach avoids the analysis perturbation sub-selection issue in the HETKF. However, it

deviates from the idea of updating the perturbations through the transform. Nevertheless, it provides an additional avenue to further reveal if the differences between the B-localization and R-localization methods for the analysis update will be dependent upon the perturbation sub-selection methods.

4.6 Experiments with the Lorenz model II

4.6.1 Lorenz model II

In this section, the Lorenz model II (Lorenz 2005; Rainwater and Hunt 2013; Fairbairn et al. 2014; Bishop et al. 2015) was used for the cycled data assimilation experiments to compare the performances of the HETKF and R-localized ETKF. Compared to the Lorenz Model I (also known as the Lorenz-96 model; Lorenz 1996), the Lorenz Model II produces spatially smoothed model trajectory. The equation for the Lorenz Model II is given by,

$$\frac{dX_n}{dt} = [X, X]_{K,n} - X_n + F, \tag{4.37}$$

where

$$[X, X]_{K,n} = \sum_{j=-J}^{J}{}' \sum_{i=-J}^{J}{}' (-X_{n-2K-i} X_{n-K-j} + X_{n-K+j-i} X_{n+K+j}) / K^2. \tag{4.38}$$

To be noted first, the usage of the symbols and letters defined in Eqs. (4.37) and (4.38) is restricted within this subsection for illustration purposes. They are not associated with the previous sections. A total of N variables are evenly distributed on a latitude cycle. Each variable X is indexed by n (n=0, 1, 2, …, N-1). The periodic boundary condition is applied. F is the forcing term. The smoothing parameter K, chosen much smaller than N, is used to define J=(K-1)/2 if K is odd and J=K/2 if K is even. The modified summation sign Σ' functions similarly as

the regular summation sign Σ except that the first and last terms are multiplied by a factor of 0.5. In Eq. (4.38), Σ' is used if K is even, otherwise, Σ' is replaced by Σ if K is odd. Following the suggestions in Lorenz (2005), the parameters of the Lorenz Model II are set as N=240, F=15 and K=8 in our experiments. The model is integrated using the fourth-order Runge-Kutta scheme. A non-dimensional time step is chosen to be 0.025 (which is equivalent to about 18 minutes in the real atmosphere).

4.6.2 Experiment design

First, the Lorenz model II was continuously integrated for a total of 80,000 time steps by selecting 240 random numbers as the initial condition. The model trajectories between the time steps of 15,001 and 30,000 serve as the simulated model climatology. The model trajectories over the last 50,000 steps are treated as the "truth" for verification. The initial ensemble is randomly drawn from the simulated model climatology. The observations are assimilated every five time steps. There are 10,000 data assimilation cycles in total. At each analysis time, the integral observations mimicking the satellite radiances are generated by averaging the "true" state variables $\mathbf{x}^t = [x_1^t, x_2^t, ..., x_n^t]^{\mathrm{T}}$ over its adjacent twenty-one grid points and adding a random noise ε drawn from a Gaussian distribution $N(0, r^2=1.32)$. The observation standard deviation r is 20% that of the simulated model climatology (Wang et al. 2007a). For example, the simulated integral observation y_i^o at the i^{th} grid point is calculated by,

$$y_i^o = \frac{1}{21} \sum_{j=i-10}^{j=i+10} x_j^t + \varepsilon. \tag{4.39}$$

In the experiments with the stochastic perturbation sub-selection method ("MP-S" and "R-S"), the K sets of perturbed observations are further generated by adding random noises drawn from the same Gaussian distribution $N(0,r^2=1.32)$ to the simulated observations. Figure 4.4 shows an example of the "true" state, non-perturbed observations and background ensemble in the first data assimilation cycle.

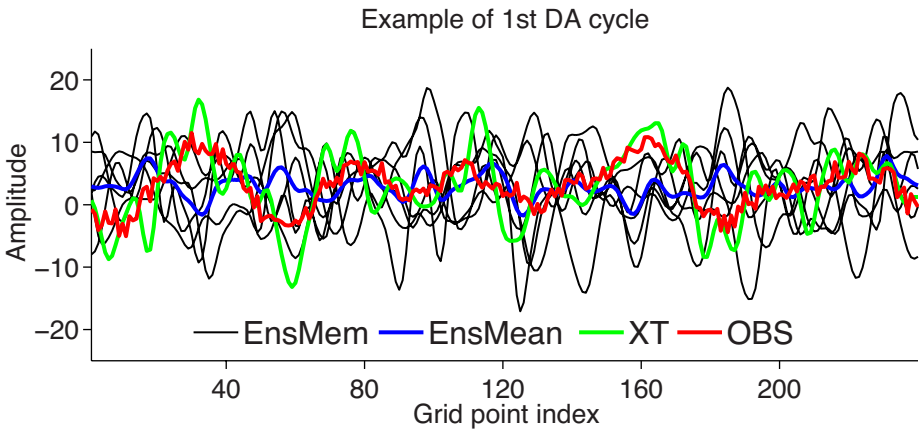

Figure 4.4 An example of the 6-member ensemble (black), the ensemble mean (blue), the simulated unperturbed observations (red) and the "true" state (green) in the first data assimilation cycle.

Furthermore, two sets of experiments were designed to more thoroughly evaluate their performances. The first set, termed as K6PX (X=30, 60, 120 and 240), runs a 6-member ensemble (K=6) but assimilates an increasing number of observations (p=30, 60, 120 and 240 correspondingly). The other set, termed as KYP240 (Y=3, 6 and 9), assimilates a total of 240 observations (p=240) but runs ensembles with an increasing size (K=3, 6 and 9). In both sets of experiments, a range of localization and inflation factors are tuned for the cycled DA experiments. Specifically, the degree of localization is determined by the parameter d in Eq. (4.9). A larger d results in stronger localization. The inflation is realized by multiplying the analysis perturbations with a factor larger than 1.0 before continuing to the next DA cycle. The

root mean square error (RMSE) between the analysis and the "truth" is calculated and averaged from the last 8,000 cycles to quantify the analysis accuracy. The percentage of the RMSE reduction (PRR) of the MP-localization method over the R-localization method is further defined as,

$$PRR = \frac{\text{RMSE(R - loc)} - \text{RMSE(MP - loc)}}{\text{RMSE(R - loc)}} \times 100\%. \tag{4.40}$$

To eliminate the random seed effects, all the experiments were repeated with eight trials by selecting different sets of random seeds (Bishop et al. 2015; Janjić et al. 2011).

4.6.3 Experiment results

1) Sensitivity of the four filters to localization and inflation factors

To obtain the minimum analysis error, extensive tuning tests were performed for each of the eight trials in each filter by combining different sets of localization and inflation factors. Figure 4.5 shows the analysis RMSE of "R-D", "MP-D", "R-S" and "MP-S" in K6P240 as a function of the localization and inflation factors. RMSE in each filter in Fig. 4.5 is averaged from all the eight trials. With the optimal localization and inflation factors (denoted by the red asterisk), the MP-localization method outperforms the R-localization method in both the deterministic and stochastic perturbation sub-selection methods. In addition, compared to the R-localization method, the MP-localization method shows less sensitivity to the localization and inflation factors. This feature is characterized by the broader blue areas in the MP-localization method in Figs. 4.5b,d. For a given localization method, compared to the stochastic perturbation sub-selection method, the deterministic perturbation sub-selection method achieves smaller minimum analysis error with less localization and inflation. The less accurate analysis in the

stochastic perturbation sub-selection method could owe to the additional sampling errors from perturbed observations (Whitaker and Hamill 2002). Overall, the MP-localization method using the deterministic perturbation sub-selection method ("MP-D") shows the most accurate analysis in K6P240.

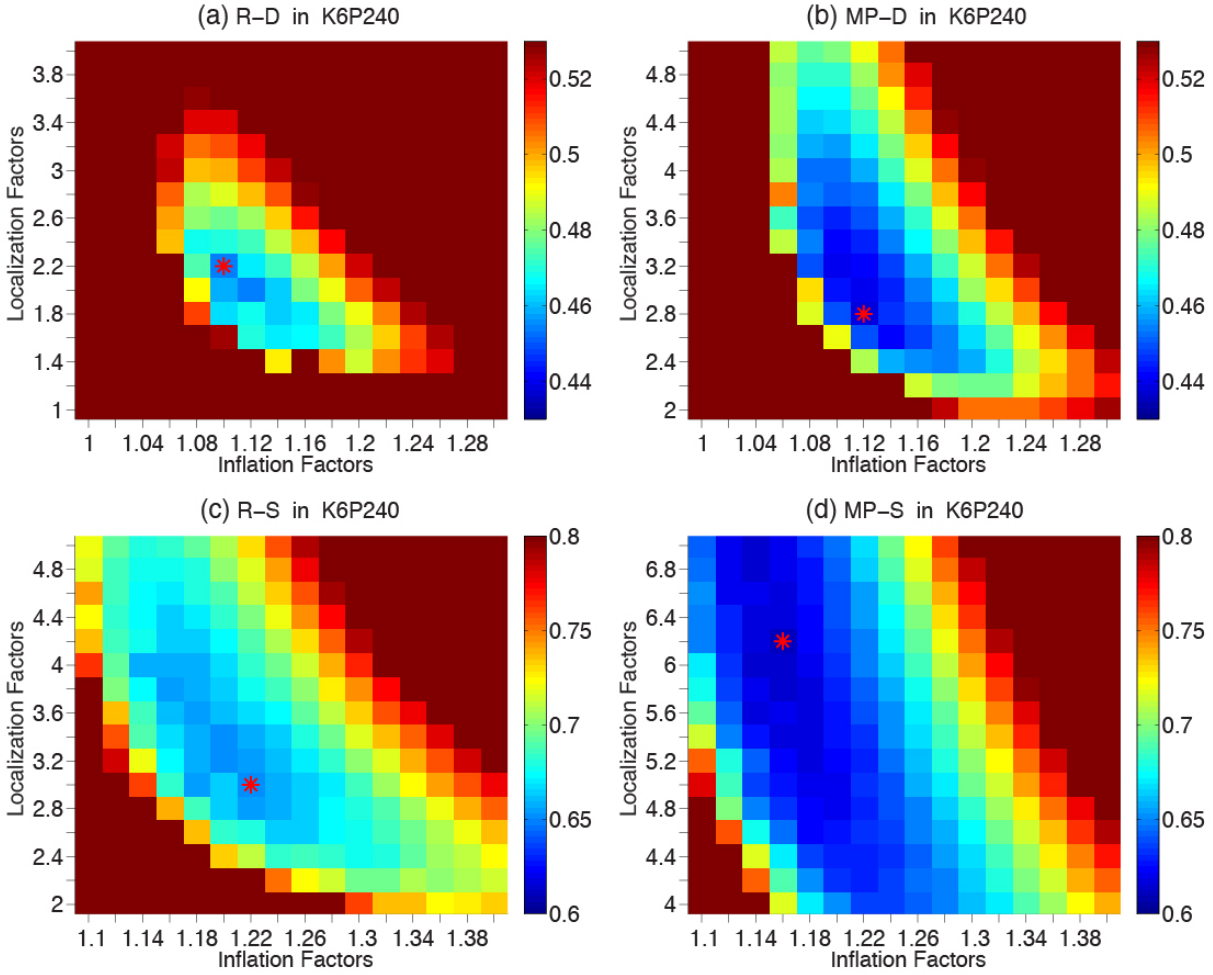

Figure 4.5 Analysis RMSE averaged over all the eight trials in the experiment of K6P240 as a function of the localization factors on the vertical axis and the inflation factors on the horizontal axis for (a) "R-D", (b) "MP-D", (c) "R-S" and (d) "MP-S". Note different color scales are applied in (a)(b) and (c)(d). Red asterisk indicates the optimal combination of the localization and inflation factors that gives the minimum analysis RMSE in each filter.

2) Filter performance as a function of the observation number

In this subsection, the K6PX (X=30, 60, 120 and 240) experiments are examined. The top panel in Fig. 4.6 shows the minimum analysis RMSE calculated from the optimal combination of the localization and inflation factors for each of the eight trials of "R-D", "MP-D", "R-S" and "MP-S". For both the deterministic and stochastic perturbation sub-selection methods, the MP-localization method significantly outperforms the R-localization method in all the four experiments. In general, the percentage of the RMSE reduction of the MP-localization method over the R-localization method tends to be slightly reduced with an increasing number of the observations (Fig. 4.8a). This is likely due to the overall improved analysis through the cycled assimilation of a larger number of observations. In the non-cycled experiments, the percentage of the RMSE reduction of the MP-localization method over the R-localization method increases with an increasing number of the observations (not shown). The latter result is more consistent with the expectation that the superiority associated with the higher rank in the estimated background error covariances becomes more pronounced in the assimilation of a larger number of the observations. For a given localization method, the deterministic perturbation sub-selection method shows smaller minimum analysis error compared to the stochastic perturbation sub-selection method.

Following section 3, the matrix $\frac{1}{w^2}\sum_{j=1}^{n}\mathbf{g}_j\mathbf{g}_j^{\mathrm{T}}$ in Eq. (4.22) for "MP-D" and "MP-S", and the matrix $\mathbf{g}_i\mathbf{g}_i^{\mathrm{T}}$ in Eq. (4.23) for "R-D" and "R-S" are calculated at the 120$^{\text{th}}$ grid point to reveal their effective localization distances (shown in the bottom panel of Fig. 4.6). Here, these two matrices are calculated by applying the optimal localization factors. In both the deterministic and stochastic perturbation sub-selection methods, the optimal effective localization distance in the

MP-localization method is tighter than that in the R-localization method. But this does not mean fewer observations will influence the analysis in the B-/MP-localization method than that in the R-localization method as discussed in section 4.3. For a given localization method, the stochastic perturbation sub-selection method requires stronger localization than the deterministic perturbation sub-selection method to obtain the minimum analysis error. As discussed in the last subsection, this could be caused by the additional sampling errors from the perturbed observations in the stochastic perturbation sub-selection method (Whitaker and Hamill 2002). The bottom panel of Fig. 4.6 shows that the optimal effective localization distance in each of "R-D", "MP-D", "R-S" and "MP-S" is not very sensitive to an increasing number of the observations. This result may be associated with the assimilation of the integral observations in our cycled data assimilation experiments. As shown in Eq. (4.39), the integral observations that are close could contain highly correlated information. Increasing the observation densities may not efficiently add additional degrees of freedom of the information provided by the integral observations. The effective localization distance is therefore likely less sensitive to an increasing number of the integral observations as designed in our cycled data assimilation experiments.

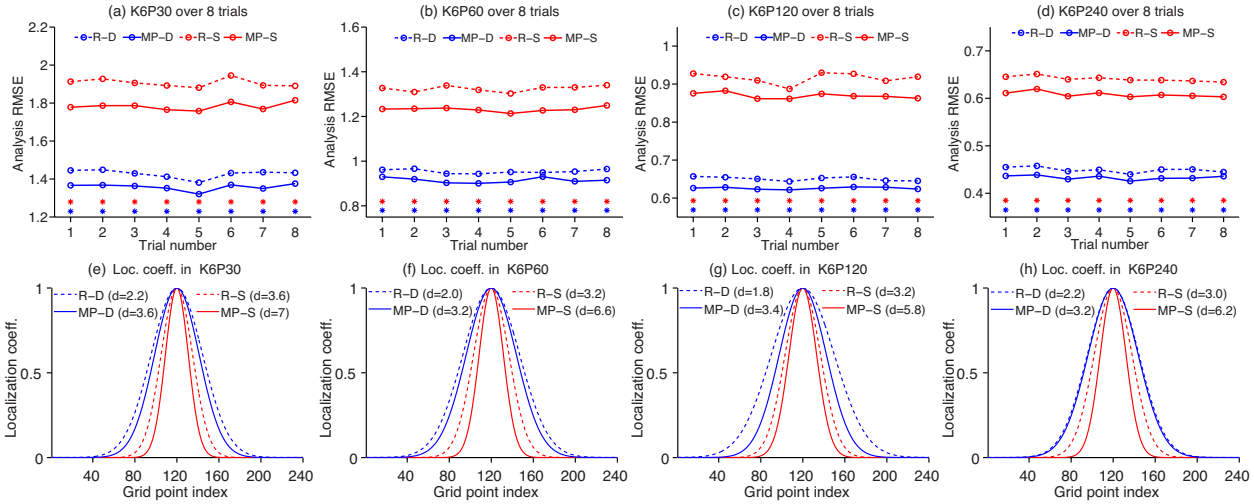

Figure 4.6 (a)-(d) Minimum analysis RMSE with the optimal combination of localization and inflation factors for "R-D" (blue dashed), "MP-D" (blue solid), "R-S" (red dashed) and "MP-S" (red solid), over the eight trials labeled on the horizontal axis in the experiments of (a) K6P30, (b) K6P60, (c) K6P120 and (d) K6P240. (e)-(h) Localization coefficients corresponding to the 120th column of the matrix $\frac{1}{w^2}\sum_{j=1}^{n}\mathbf{g}_j\mathbf{g}_j^{\mathrm{T}}$ as shown in Eq. (4.22) for "MP-D" (blue solid) and "MP-S" (red solid) and of the matrix $\mathbf{g}_i\mathbf{g}_i^{\mathrm{T}}$ for as shown in Eq. (4.23) for "R-D" (blue dashed) and "R-S" (red dashed) by applying the optimal tuned localization factor *d* defined in Eq. (9) and listed at the top of the figures of (e) K6P30, (f) K6P60, (g) K6P120 and (h) K6P240.

3) Filter performance as a function of the ensemble size

Figure 4.7 shows the results for the KYP240 (Y=3 and 9) experiments. When the ensemble size is reduced from 6 to 3, the relative performance of the MP-localization and R-localization methods (Fig. 4.7a) is similar to that of K6P240. Specifically, in both the deterministic and stochastic perturbation sub-selection methods, the MP-localization method significantly outperforms the R-localization method. For a given localization method, the deterministic perturbation sub-selection method shows smaller minimum analysis error than the stochastic perturbation sub-selection method. In addition, in K3P240, the MP-localization method using the stochastic perturbation sub-selection method ("MP-S") even shows more accurate analysis than the R-localization method using the deterministic perturbation sub-selection method ("R-D").

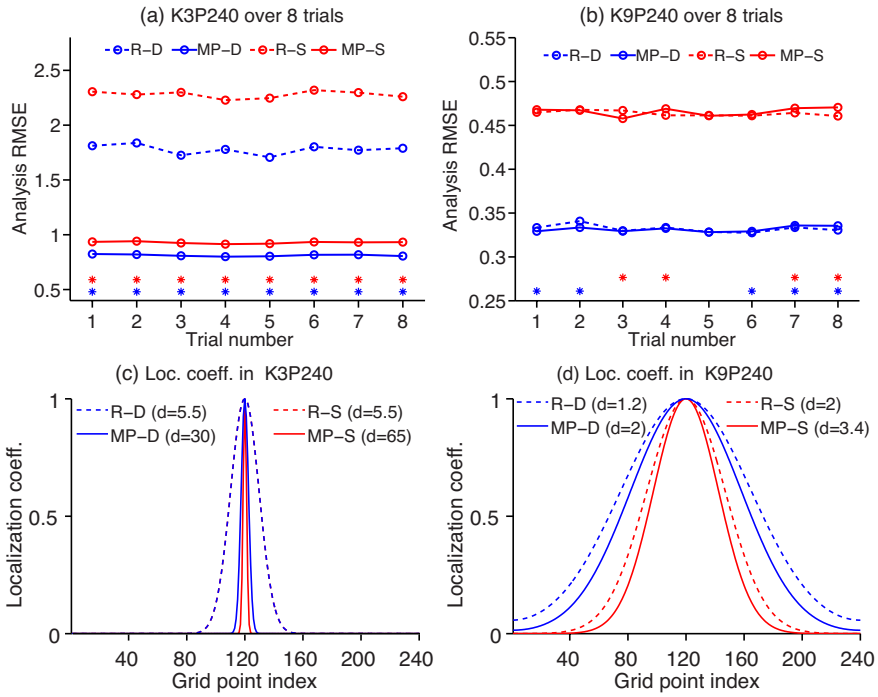

Figure 4.7 Same as Figure 4.6 except for the experiments of (a)(c) K3P240 and (b)(d) K9P240.

By increasing the ensemble size to 9 (K9P240), for both perturbation sub-selection methods, the MP-localization and R-localization methods perform comparably and their difference is statistically insignificant in most of the eight trials. Consistently, the percentage of the RMSE reduction of the MP-localization method over the R-localization method is reduced with an increasing ensemble size in both perturbation sub-selection methods (Fig. 4.8b). This is within the expectation that the higher rank from the B-/MP-localization method would contribute more positively to alleviating the rank deficiency issue and thus improving the analysis for a small ensemble. This may further suggest that the improved analysis in the B-/MP-localization method is likely associated with its higher rank as demonstrated in section 4.3.

The effective localization distances calculated from the optimal localization factors are shown for the KYP240 experiments in the bottom panel of Fig. 4.7. For a given perturbation sub-

selection method, the optimal effective localization distance in the MP-localization method is tighter than that of the R-localization method. For both localization methods, the optimal effective localization distance from the stochastic perturbation sub-selection method is tighter than the deterministic perturbation sub-selection method except for the K3P240 experiment with the R-localization method. Figure 4.7 also demonstrates as expected that the effective localization distances in each of "R-D", "MP-D", "R-S" and "MP-S" becomes wider with a larger ensemble size.

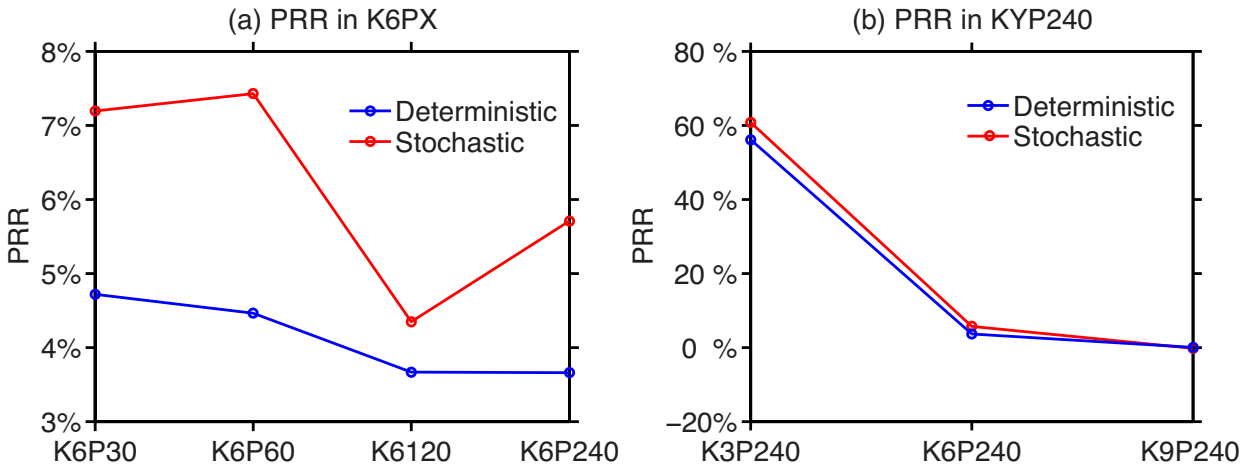

Figure 4.8 Percentage of the RMSE reduction of the MP-localization method over the R-localization method in the deterministic (blue) and stochastic (red) perturbation sub-selection methods in the experiments of (a) K6PX and (b) KYP240.

4.7 Conclusion and discussion

A mathematical demonstration is first provided to compare the B-localization and R-localization methods. It is shown that when the same effective localization function is applied, the B-localization method achieves a higher rank than the R-localization method in the localized background error covariance matrix. The mathematical demonstration is further illustrated and validated using a simple example. Further examination suggests that all the observations will

contribute to updating a single gird point in the B-localization method. However, the analysis at a particular grid point in the R-localization method is influenced by limited observations that are close. Meanwhile, the mathematical demonstration also shows that the B-localization method can be realized through extending and modulating the raw background ensemble perturbations or the MP-localization method. Specifically, in the MP-localization method, each raw ensemble perturbation vector is modulated through an element-wise multiplication with each of the modulation functions. To improve the computational efficiency, the modulation functions are calculated from the leading eigenvalues and eigenvectors of the original B-localization matrix. The resulting MP-localized background error covariance matrix is thus consistent with that applying the traditional B-localization method. In the mathematical demonstration, it proves that the R-localization method can also be expressed in the form of the modulated ensemble perturbations as in the B-localization method. The B-/MP-localization method is then implemented in the ETKF and further compared with the R-localization method using the same ETKF algorithm. Due to the higher rank from the B-localization method as derived in the mathematical demonstration, the B-/MP-localized ETKF is termed as the high-rank ETKF (HETKF) to distinguish it from the classic R-localized ETKF.

Extensive cycled data assimilation experiments were conducted to compare the performances of the HETKF and R-localized ETKF using the Lorenz model II. Using the same ETKF algorithm warrants a homogeneous comparison between these two localization methods, so that it is more straightforward to relate their resulting analysis performances with the localization differences. The results show that the HETKF significantly and consistently improves the analysis over the R-localized ETKF especially for a small ensemble. Since the

higher rank from the HETKF is expected to contribute more positively to mitigating the rank deficiency issue for a small ensemble, the improved analysis of the HETKF over the R-localized ETKF is likely associated with the higher rank from the B-/MP-localization method. In addition, the advantage of the HETKF over the R-localized ETKF tends to be slightly reduced with the increasing number of the observations. This result could be attributed to the improved accuracy of the system through the cycled assimilation of a larger number of observations. Furthermore, the HETKF is less sensitive to the localization length scales and inflation factors than the R-localized ETKF. In all the experiments, the HETKF shows tighter optimal effective localization distance than the R-localized ETKF. The above conclusion of comparing the HETKF and R-localized ETKF does not rely on the perturbation sub-selection methods in the HETKF.

It is also found that in both the HETKF and R-localized ETKF, the stochastic perturbation sub-selection method shows larger analysis error than the deterministic perturbation sub-selection method. In addition, in both filters, the stochastic perturbation sub-selection method generally requires stronger localization and larger inflation than the deterministic perturbation sub-selection method to obtain the minimum analysis error except for the experiment with the R-localized ETKF and a very small ensemble (e.g., K3P240). This can be attributed to the sampling errors by perturbing the observations in the stochastic perturbation sub-selection method (Whitaker and Hamill 2002).

In this study, the improved analysis from the B-/MP-localization method over the R-localization method is demonstrated using the same ETKF algorithm in the Lorenz model II. This is consistent with the results in Janjić et al. (2011) and Nerger et al. (2012) that adopted different EnKF variants for comparison. To implement the HETKF in the operational modeling

systems, additional treatments are likely needed for the computational concerns. For example, a parallel, patch-based implementation like the LETKF can be adopted to improve the computational scalability. Further diagnostics (not shown) showing the analysis errors calculated from the full 10,000 cycles (i.e., including those cycles before the errors get stabilized) indicate that the HETKF requires less time to converge compared to the R-localized ETKF. This feature and the less sensitivity of the HETKF to the localization length scales and inflation factors are attractive for the operational model applications.

Chapter 5: A Simultaneous Multi-scale Data Assimilation using Scale Dependent Localization in GSI-based Hybrid 4DEnVar for NCEP FV3-based GFS

5.1. Introduction

Ensemble-based data assimilation (DA) approach, such as ensemble Kalman filter (EnKF; Evensen 1994) and hybrid ensemble-variational (EnVar) DA (Hamill et al. 2000; Lorenc 2003; Buehner 2005; Wang et al. 2007; Wang 2010), has been widely adopted in many operational numerical weather prediction (NWP) centers to produce initial conditions for medium-range forecasts. In the ensemble-based DA approach, an ensemble of short-range forecasts estimate flow-dependent background error covariances. This contrasts to the traditional variational DA approach that assumes static background error covariances. The advantage of the ensemble-based DA approach over the pure variational DA approach has been demonstrated in the global and regional applications (Wang et al. 2007a, 2008ab, 2013; Wang 2011; Buehner et al. 2013; Clayton et al. 2013; Gustafsson et al. 2014; Wang and Lei 2014; Lorenc et al. 2015; Kleist and Ide 2015a,b; Kutty and Wang 2015; Buehner et al. 2015).

Limited computational resources constrain the affordable ensemble size to be much smaller than degrees of freedom of the model itself (Houtekamer and Zhang 2016). This results in sampling error in the ensemble-based DA approach. Its typical features are distant spurious correlations. Successful application of ensemble-based DA approach relies on efficient treatment of sampling error. Directly increasing ensemble size will reduce sampling error (Miyoshi et al. 2014; Lei and Whitaker 2017; Huang and Wang 2018). But it can be computationally prohibitive

especially for operational NWP applications. Alternatively, covariance localization is commonly applied in the ensemble-based DA approach to gradually attenuate and even eliminate distant spurious correlations caused by sampling error (Houtekamer and Mitchell 2001, 2005; Huang et al. 2019). Applying covariance localization contributes to improved analyses and subsequent forecasts (Houtekamer and Mitchell 1998, 2001; Bishop and Hodyss 2009; Buehner 2012; Anderson and Lei 2013; Gasperoni and Wang 2015).

Rapid advancement of high-performance computing allows future global NWP model to resolve much wider range of scales. DA that appropriately updates a wide range of scales will be required. Zhang et al. (2009) performed a multi-step sequential DA update by separately assimilating different groups of observations with applying different localization length scales. But this multi-step sequential DA update could potentially lose some useful information that may exist in a simultaneous assimilation of all available observations (Caron and Buehner 2018). Miyoshi and Kondo (2013) combined two sets of independent analysis increments from assimilating the same set of observations. Each set applied different amount of localization. Both methods showed improved analyses and subsequent forecasts in the EnKF systems compared to applying fixed uniform localization once at all scales.

While the aforementioned methods took multiple-steps or adopted sequential-update, a single-step simultaneous multi-scale update was proposed recently by introducing scale-dependent localization (SDL) in the EnVar framework (Buehner 2012; Buehner and Shlyaeva 2015). This simultaneous SDL can be classified into two variants. The first variant completely eliminates the cross-waveband covariances (Buehner 2012) (hereafter, referred to as SDL-NoCross). Mathematically, SDL-NoCross equivalently applies a local spatial averaging of

ensemble covariances which may alleviate sampling error and improve the accuracy of ensemble covariances especially for a small ensemble (Buehner and Charron 2007). Compared to applying fixed uniform localization once at all scales, SDL-NoCross improved the general global forecast skills in a global 3DEnVar (Buehner 2012) and 4DEnVar (Lorenc 2017) system.

The second simultaneous SDL variant takes into account cross-waveband covariances (hereafter, referred to as SDL-Cross) (Buehner and Shlyaeva 2015). Compared to SDL-NoCross, SDL-Cross may retain more heterogeneity of error covariances (Caron and Buehner 2018). SDL-Cross was demonstrated in a regional 3DEnVar sea-ice DA system to perform better than fixed unfiorm localization at all scales (Buehner and Shlyaeva 2015). Caron and Buehner (2018) implemented SDL-Cross in a global 3DEnVar system, and found improved global forecasts over scale-invariant localization. Furthermore, Caron et al. (2019) comparing SDL-NoCross and SDL-Cross in a regional 3DEnVar system showed that SDL-NoCross produced more accurate forecasts than SDL-Cross using a 25-member ensemble, while both performed comparably when using a 75-member ensemble populated by time-lagged method (Lorenc 2017; Huang and Wang 2019). Caron et al. (2019) further hypothesized that the relative performances between SDL-NoCross and SDL-Cross could be associated with the accuracy of the estimated cross-waveband covariances in SDL-Cross that depended on ensemble size.

This study addresses several additional questions on simultaneous SDL using the US NWS GSI-based hybrid 4DEnVar system (Wang and Lei 2014; Kleist and Ide 2015a) . The GSI-based 4DEnVar system was recently integrated with the US next generation non-hydrostatic Finite-Volume Cubed-sphere dynamical core (FV3)-based GFS model (Chen et al. 2019; Zhou et al. 2019). To achieve the goal of exploring new scientific questions associated with simultaneous

SDL, we first demonstrated both SDL methods mathematically in a full B-preconditioned EnVar framework, and implemented both in the GSI-based hybrid 4DEnVar system. SDL formulations within an EnVar framework preconditioned on the full B and square-root of B were also discussed in Caron et al. (2019). Given SDL-NoCross and SDL-Cross were only compared for regional applications previously, this study first evaluates and compares both approaches for the general global forecasts. Second, in previous studies, SDL was implemented with no explicit level-dependence and compared with level- and scale-invariant horizontal localization. In our study, the baseline operational GSI-based hybrid 4DEnVar system for the FV3-based GFS applies scale-invariant, but level-dependent horizontal localization. Therefore, how does SDL perform relative to the more strict baseline which adopts scale-invariant, but level-dependent horizontal localization? Third, tropical storm track forecasts were compared between the GSI-based 3DEnVar and 4DEnVar that adopted the level-dependent scale-invariant horizontal localization (Wang and Lei 2014). The tropical storm track is influenced by the general large-scale environmental flow (Zong and Wu 2015). Given our cycled experiment configuration in section 5.3 that uses an ensemble resolution about 50 km, it is interesting to investigate how SDL-NoCross and SDL-Cross would influence the tropical storm track forecasts compared to the scale-invariant localization in the current GSI-based 4DEnVar system. Fourth, efficient scale separation in SDL is essential and remains to be investigated. How does the performance of SDL vary with different numbers of decomposed wavebands (e.g., two versus three)? Finally, diagnostics were performed to understand different performances between scale-invariant localization, SDL-NoCross and SDL-Cross.

This chapter is organized as follows. Section 5.2 describes the SDL formulation and

implementation in the GSI-based hybrid 4DEnVar system. Experiment design is described in section 5.3. Sections 5.4 and 5.5 discuss the experiment results. In section 5.6, some diagnostics are presented to understand the results in sections 5.4 and 5.5. The computational cost is compared in section 5.7. Section 5.8 presents conclusion and discussion.

5.2 SDL formulation and implementation in the GSI-based 4DEnVar system

5.2.1 General SDL formulation in the GSI-based 4DEnVar system

The GSI-based 4DEnVar system is formulated and implemented based on extended control variable method, to incorporate the ensemble background covariances within the traditional variational framework (Wang 2010b; Wang et al. 2013; Wang and Lei 2014). Mathematically, it is equivalent to linearly combining the static and ensemble background covariances (Wang et al. 2007b, 2008a). In this study, SDL was implemented in the GSI-based hybrid 4DEnVar system by further extending control variables. In this section, the general SDL formulation in the 4DEnVar system is first illustrated following the notations in Wang et al. (2013) and Wang and Lei (2014). Specific implementations of SDL-NoCross and SDL-Cross are then detailed. To highlight the variables of further extended dimension due to applying SDL, a "hat" sign is labeled above the letter or symbol.

Following Buehner and Shlyaeva (2015), the normalized ensemble perturbations are decomposed into a set of J overlapping wavebands or scales which are indexed by j=1, …, J,

$$\mathbf{x}_{k,j}^{e} = \mathbf{\Psi}_{j}\mathbf{x}_{k}^{e} \tag{5.1}$$

where $\mathbf{x}_{k}^{e}$ denotes the kth ensemble perturbation vector normalized by $(K\text{-}1)^{1/2}$, K is the ensemble size, $\mathbf{\Psi}_{j}$ is the spectral filter function that extracts the jth waveband, and $\mathbf{x}_{k,j}^{e}$ is the kth

normalized ensemble perturbation vector that only contains the *j*th waveband.

In the 4DEnVar with applying SDL, the analysis increment $\mathbf{x}'_t$ at time *t* in a DA window is calculated as

$$\mathbf{x}'_t = \mathbf{x}'_1 + \hat{\mathbf{I}} \sum_{k=1}^{K} \left[\hat{\mathbf{a}}_k \circ (\hat{\mathbf{x}}_k^e)_t \right], \tag{5.2}$$

where

$$\hat{\mathbf{I}} = \left[\begin{array}{cccc} \mathbf{I} & \mathbf{I} & \cdots & \mathbf{I} \end{array} \right], \tag{5.3}$$

and

$$\hat{\mathbf{a}}_k = \left[\begin{array}{c} \mathbf{a}_{k,1} \\ \mathbf{a}_{k,2} \\ \vdots \\ \mathbf{a}_{k,J} \end{array} \right], \text{ and } \quad \hat{\mathbf{x}}_k^e = \left[\begin{array}{c} \mathbf{x}_{k,1}^e \\ \mathbf{x}_{k,2}^e \\ \vdots \\ \mathbf{x}_{k,J}^e \end{array} \right]. \tag{5.4}$$

The first term on the right-hand side of Eq. (5.2) is the analysis increment associated with the static background error covariances. $\hat{\mathbf{I}}$ contains *J* identity matrices aligned in a row. $\hat{\mathbf{x}}_k^e$ is a vector that concatenates *J* vectors of decomposed $\mathbf{x}_{k,j}^e$ (*j*=1, …, *J*) for the *k*th member. $\mathbf{a}_{k,j}$ is a three dimensional vector that corresponds to the control variable vector at the *j*th waveband for the *k*th member, and $\hat{\mathbf{a}}_k$ is further extended control variable vector that concatenates *J* vectors of $\mathbf{a}_{k,j}$ for the *k*th member. The sign "o" denotes the Schur product. Compared to Wang and Lei (2014) without applying SDL, the control variable vector $\hat{\mathbf{a}}_k$ in Eq. (5.4) for the *k*th member varies with the waveband index *j*. Its dimension is increased by *J*. As in Wang and Lei (2014),

the same set of $\hat{\mathbf{a}}_k$ is applied for all control variables that include surface pressure, virtual temperature, wind, relative humidity, cloud water mixing ratio and ozone mixing ratio at different time levels.

The analysis increment in Eq. (5.2) is obtained by minimizing the following cost function:

$$J(\mathbf{x}_1', \hat{\mathbf{a}}) = \frac{1}{2}\beta_1(\mathbf{x}_1')^{\mathrm{T}}\mathbf{B}_1^{-1}(\mathbf{x}_1') + \frac{1}{2}\beta_2(\hat{\mathbf{a}})^{\mathrm{T}}\hat{\mathbf{A}}^{-1}(\hat{\mathbf{a}}) + \frac{1}{2}\sum_{t=1}^{L}(\mathbf{y}_t^{o\prime} - \mathbf{H}_t\mathbf{x}_t')^{\mathrm{T}}\mathbf{R}_t^{-1}(\mathbf{y}_t^{o\prime} - \mathbf{H}_t\mathbf{x}_t'), \tag{5.5}$$

On the right-hand side of Eq. (5.5), the first term is associated with the static background covariances $\mathbf{B}_1$. In the second term, $\hat{\mathbf{a}}$ is the extended control variable vector that concatenates K vectors of $\hat{\mathbf{a}}_k$ in Eq. (5.4). $\hat{\mathbf{A}}$ is a block-diagonal matrix that defines the localization matrix for within- and cross-waveband ensemble covariances (see more details later). In the third term, $\mathbf{y}_t^{o\prime}$, $\mathbf{H}_t$ and $\mathbf{R}_t$ are the observation innovation vector, linearized observation operator matrix and observation covariance matrix at time t, respectively. L is the number of time levels spanning the DA window (e.g., 6 hours). In addition, parameters β_1 and β_2 control the weights of the static and ensemble background covariances and $(1/\beta_1)+(1/\beta_2)=1$ is required as in Wang et al. (2007b, 2008a).

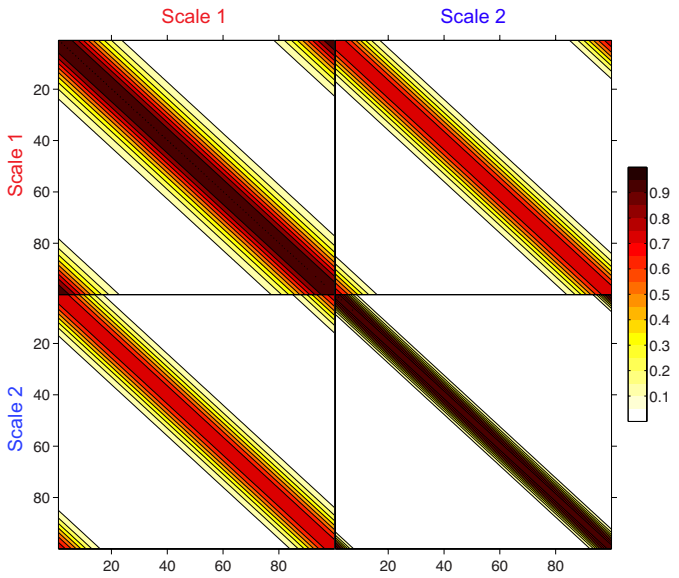

Figure 5.1 Illustration of the scale-dependent spatial localization matrix between Scale 1 and Scale 2 that represent large and small scales, respectively, using a one-dimensional periodic domain of 100 grid points.

In the global GSI-based 4DEnVar system, the localization defined in $\hat{\mathbf{A}}$ is realized through spectral filter transformation on the horizontal direction and recursive filter transformation on the vertical direction. More details of implementing the horizontal and vertical localization were described in Wang (2010) and Wang et al. (2013). In this study, SDL is only applied for horizontal localization. Specifically, the explicit formula of $\hat{\mathbf{A}}$ can be written as

$$\hat{\mathbf{A}} = \begin{bmatrix} \hat{\mathbf{L}} & & & \\ & \hat{\mathbf{L}} & & \\ & & \ddots & \\ & & & \hat{\mathbf{L}} \end{bmatrix} \tag{5.6}$$

Each of the K blocks in $\hat{\mathbf{A}}$ contains the same predefined localization matrix $\hat{\mathbf{L}}$ with unit diagonal elements following Buehner and Shlyaeva (2015),

$$\hat{\mathbf{L}} = \begin{bmatrix} \hat{\mathbf{L}}_{1,1} & \hat{\mathbf{L}}_{1,2} & \cdots & \hat{\mathbf{L}}_{1,J} \\ \hat{\mathbf{L}}_{2,1} & \hat{\mathbf{L}}_{2,2} & \cdots & \hat{\mathbf{L}}_{2,J} \\ \vdots & \vdots & \ddots & \vdots \\ \hat{\mathbf{L}}_{J,1} & \hat{\mathbf{L}}_{J,2} & \cdots & \hat{\mathbf{L}}_{J,J} \end{bmatrix} = \begin{bmatrix} \hat{\mathbf{L}}_{1,1}^{1/2} \\ \hat{\mathbf{L}}_{2,2}^{1/2} \\ \vdots \\ \hat{\mathbf{L}}_{J,J}^{1/2} \end{bmatrix} \begin{bmatrix} \hat{\mathbf{L}}_{1,1}^{\mathrm{T}/2} & \hat{\mathbf{L}}_{2,2}^{\mathrm{T}/2} & \cdots & \hat{\mathbf{L}}_{J,J}^{\mathrm{T}/2} \end{bmatrix} \tag{5.7}$$

where $\hat{\mathbf{L}}_{j1,j2} = \hat{\mathbf{L}}_{j1,j1}^{1/2} \hat{\mathbf{L}}_{j2,j2}^{\mathrm{T}/2}$ ($j1$=1, …, J and $j2$=1, …, J) defines the localization matrix for the ensemble covariances between the $j1$th and $j2$th wavebands. Eq. (5.7) ensures the complete localization matrix is positive semi-definite (Buehner and Shalyeva 2015). Given the design in Eq. (5.7), the within-waveband localization matrices have unit diagonal elements, while the cross-waveband localization matrices display less-than-one diagonal elements (Fig. 5.1).

5.2.2 Specific implementation of SDL-NoCross and SDL-Cross in the GSI-based 4DEnVar system

In this subsection, implementations of SDL-NoCross and SDL-Cross in the GSI-based 4DEnVar system are further described. Two major implementation differences are involved. One is the definition of the spectral filter function $\mathbf{\Psi}_j$ in Eq. (5.1). The other is whether or not to zero out the cross-waveband localization matrix $\hat{\mathbf{L}}_{j1,j2}$ ($j1 \neq j2$) in Eq. (5.7).

In SDL-Cross, it requires that the spectral filter functions $\mathbf{\Psi}_j$ over J wavebands sum to one to recover the original raw ensemble perturbations from their decomposed components (Buehner and Shlyeva 2015) , and that the cross-waveband localization matrix $\hat{\mathbf{L}}_{j1,j2}$ ($j1 \neq j2$) is retained to partially include cross-waveband covariances. In a particular scenario of applying the same amount of localization at different wavebands, SDL-Cross is equivalent to applying fixed

uniform localization once at all scales (Buehner and Shlyaeva 2015).

In contrast, to implement SDL-NoCross, it requires that the squared spectral filter functions Ψ_j over J wavebands sum to one to maintain the total raw ensemble variances (Buehner 2012), and that the cross-waveband localization matrix $\hat{\mathbf{L}}_{j1,j2}$ ($j1 \neq j2$) is set to be zero to completely remove cross-waveband covariances.

5.3 Experiment design

The GSI-based 4DEnVar DA system and the FV3-based GFS model were used for one-month cycled DA experiments from 0000 UTC 25 August to 1800 UTC 24 September 2017. More details about the GSI-based EnVar system can be found in Wang et al. (2013) and Wang and Lei (2014). The assimilated observations over a six-hour DA window include all the conventional and satellite observations in the operational NCEP global DA system. Satellite radiance data assimilation applied the same observation quality control and bias correction in the operational global DA system (Zhu et al. 2014).

The baseline 4DEnVar experiment (W1-Ope in Table 5.1) was set up similarly as the operational system using a dual-resolution configuration. One-member control background is at a resolution of C384 (~ 25 km), while 80-member ensemble background is at a reduced resolution of C192 (~ 50 km) (Lin 2004; Harris and Lin 2013). In the DA update, the one-member control background was updated using the 4DEnVar algorithm (Wang and Lei 2014; Kleist and Ide 2015a). Specifically, 12.5% static and 87.5% ensemble background covariances were combined to construct its hybrid form as in the operational system. Both percentage numbers correspond to $(1/\beta_1)$ and $(1/\beta_2)$ in section 5.2.1, respectively. Three-hourly ensemble perturbations were

ingested in the 4DEnVar update to account for the temporal evolution. To deal with sampling error, localization was applied on both horizontal and vertical directions. As in the operational system, the horizontal localization length scale varies with model level (e.g., black dotted curve in Fig. 5.2), and a fixed uniform vertical localization length scale is applied at all model levels (Table 5.1). In addition, a tangent linear normal mode initialization constraint (TLNMC, Kleist et al. 2009) was applied to alleviate the imbalance in the control analysis as in the operational system.

Table 5.1 List of DA experiments

Exps	**Number of wavebands**	**Horizontal localization length scale (*e*-folding distance)**	**Vertical localization length scale (scale height, *e*-folding distance)**
W1-Ope	1	Level-dependent length scale for full-scale ensemble perturbations (black curve in Fig. 5.2)	0.5
W1-1000	1	1000 km for full-scale ensemble perturbations	
W1-300	1	300 km for full-scale ensemble perturbations	
W2-NoCross	2	1000 km and 300 km for large- and small-scale ensemble perturbations, respectively	
W2-Cross	2		
W3-NoCross	3	1000 km, 650 km and 300 km for large-, medium- and small-scale ensemble perturbations, respectively	
W3-Cross	3		

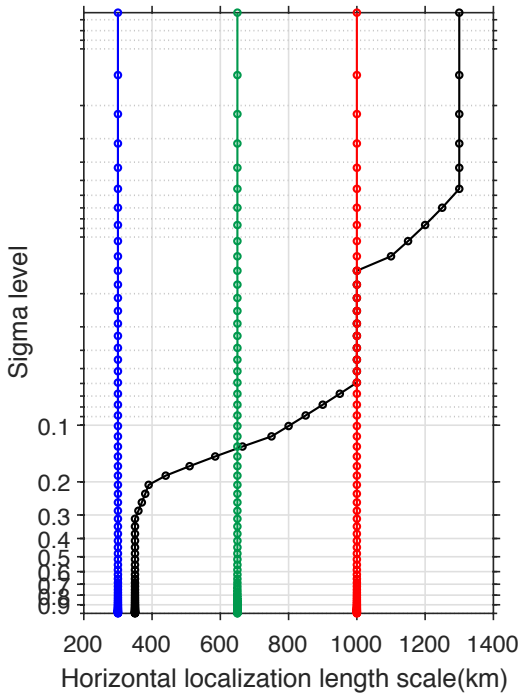

Figure 5.2 Level-dependent horizontal localization length scales (black) for W1-Ope and level-invariant horizontal localization length scales at large- (red), medium- (green) and small- (blue) scale wavebands in the two- and three-waveband SDL experiments in Table 5.1. Note that the horizontal localization length scale is *e*-folding distance.

The 4D local ensemble transform Kalman filter (LETKF, Hunt et al. 2007) was adopted to update the 80-member background ensemble. In the LETKF, the observation operators were calculated through the GSI. To remedy sampling error, the localization was applied in the LETKF and defined by the Gaspari-Cohn function (Gaspari and Cohn 1999). The exact cut-off distance in the Gaspari-Cohn function is equivalent to multiplying the *e*-folding localization length scales in the 4DEnVar update by a factor of 2.577 (Pan et al. 2014). To remedy the background ensemble spread deficiency, multiplicative inflation (Whitaker et al. 2012; Lei and Whitaker 2016, 2017) was employed to relax the posterior ensemble spread back to 85% of the prior ensemble spread. Stochastic parameterization schemes (Palmer et al. 2009; Lei and Whitaker 2016, 2017; Huang and Wang 2018) were further applied to account for model

uncertainty.

The GFS model using the non-hydrostatic Finite-Volume Cubed-sphere dynamical core (FV3, Lin 2004; Harris and Lin 2013) was used to provide the control and ensemble background forecasts. The FV3-based GFS was configured similarly as in the pre-operational tests in Phase II of the Next-Generation Global Prediction System (NGGPS) project (National Weather Service, 2020). The model configurations were detailed in Chen et al. (2019) and Zhou et al. (2019). Due to computational constraints, the experiments in this study were performed at a reduced horizontal resolution compared to the pre-operational tests. There are a total of 64 model levels in the current FV3-based GFS model. The 4D incremental analysis update (4DIAU) was further applied for both control and ensemble forecasts to improve the balance during the model integration (Lorenc et al. 2015; Lei and Whitaker 2016, 2017; Huang and Wang 2018).

In SDL, the optimal way of performing scale separation remains to be investigated. A hurricane example (Fig. 5.3a) was used to assist in scale separation in our SDL experiments. As our initial examination, the first set of SDL experiments adopted two wavebands, referred to as W2-NoCross and W2-Cross in Table 5.1 that apply SDL-NoCross and SDL-Cross, respectively. For the scale separation in the two-waveband SDL experiments, the hurricane at relatively small scale (Fig. 5.3c) was isolated from the general large-scale environmental flow (Fig. 5.3b) using the example in Fig. 5.3a. The resultant spectral filter functions (Fig. 5.4a,b) at the large- and small-scale wavebands cross each other roughly at wavelength 2500 km or at total wavenumber 16. The spectral filter functions in SDL-NoCross were defined as the square root of those in SDL-Cross following Caron et al. (2019). It is meant to satisfy that the spectral filer functions sum to one in SDL-Cross, while the squared spectral filter functions sum to one in SDL-NoCross

as discussed in section 5.2.2. To define the horizontal localization length scale in the two-waveband SDL experiments, several tuning tests were performed by referring to the horizontal localization length scales applied in Caron and Buehner (2018) and the operational level-dependent horizontal localization length scales in W1-Ope. Finally, the 1000 km and 300 km *e*-folding distances were selected as the horizontal localization length scales at all model levels for the large- and small-scale ensemble perturbations in the two-waveband SDL experiments, respectively. To provide a more homogeneous comparison with the two-waveband SDL experiments applying level-invariant horizontal localization, another two experiments of W1-1000 and W1-300 in Table 5.1 were further designed. Different from W1-Ope applying the level-dependent horizontal localization, however, W1-1000 and W1-300 apply the level-invariant horizontal localization length scales, that is, 1000 km and 300 km *e*-folding distances, respectively.

To explore how the SDL performance varies with the number of decomposed wavebands, three-waveband SDL experiments were further designed such as W3-NoCross and W3-Cross in Table 5.1 that apply SDL-NoCross and SDL-Cross, respectively. For the scale separation in the three-waveband SDL experiments, the large-scale background in the two-waveband SDL experiments was further decomposed to two wavebands, that is, the large- and medium-scale wavebands in the three-waveband SDL experiments. As a result, the mid-latitude high pressure (Fig. 5.3e) that appears in the large-scale waveband in the two-waveband SDL experiments is further isolated from the global large-scale environmental flow (Fig. 5.3d). The background at small-scale waveband in the three-waveband SDL experiments (Fig. 5.3f) was retained similarly as in the two-waveband SDL experiments (Fig. 5.3c) typically featured by the hurricane. Figures

5.4c,d show the resultant spectral filter functions at large-, medium- and small-scale wavebands in the three-waveband SDL experiments. Additional tuning tests were performed to determine the horizontal localization length scales in the three-waveband SDL experiments. But it did not provide additional benefits by further increasing the horizontal localization length scale beyond 1000 km *e*-folding distance for the large-scale waveband in the three-waveband SDL experiments. The same set of 1000 km and 300 km *e*-folding distances were thus applied at the large- and small-scale wavebands in the three-waveband SDL experiments. Additionally, the 650 km *e*-folding distance was selected at the medium-scale waveband in the three-waveband SDL experiments. Therefore, the main difference in the two- and three-waveband SDL experiment designs is tighter horizontal localization applied at medium-scale waveband in the three-waveband SDL experiments. This would facilitate to interpret performance differences between the two- and three-waveband SDL experiments.

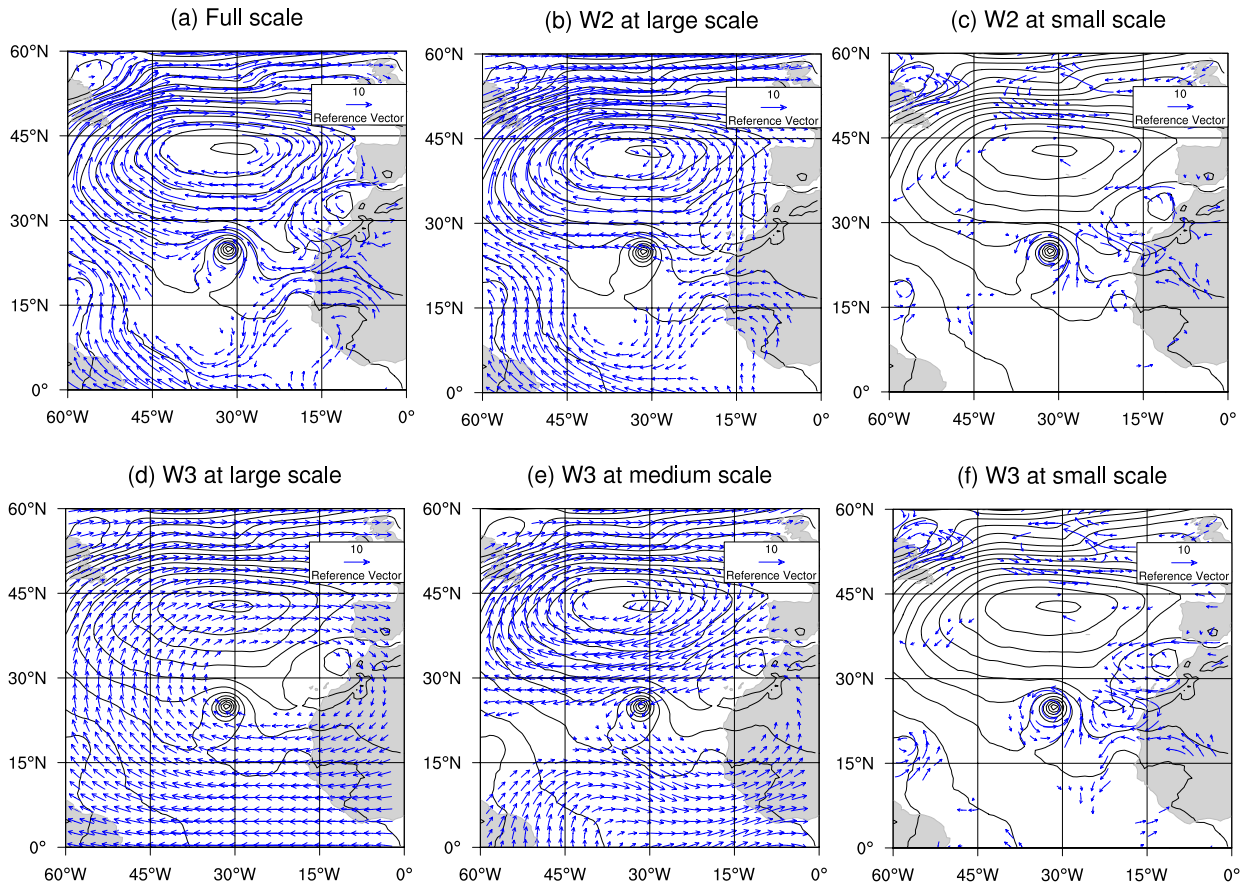

Figure 5.3 Two-dimensional wind vectors at 850 hPa in a hurricane example at (a) full-scale, and at (b) large- and (c) small-scale wavebands in the two-waveband SDL experiments, and at (d) large-, (e) medium- and (f) small-scale wavebands in the three-waveband SDL experiments. The grey contours beneath the wind vectors denote the geopotential height at 850 hPa.

Experiment descriptions are detailed in Table 5.1. For all the experiments in Table 5.1, they applied the 0.5 scale-height *e*-folding distance at all model levels for the vertical localization. On the other hand, in the LETKF update, all the experiments applied the same horizontal and vertical localization length scales as in W1-Ope.

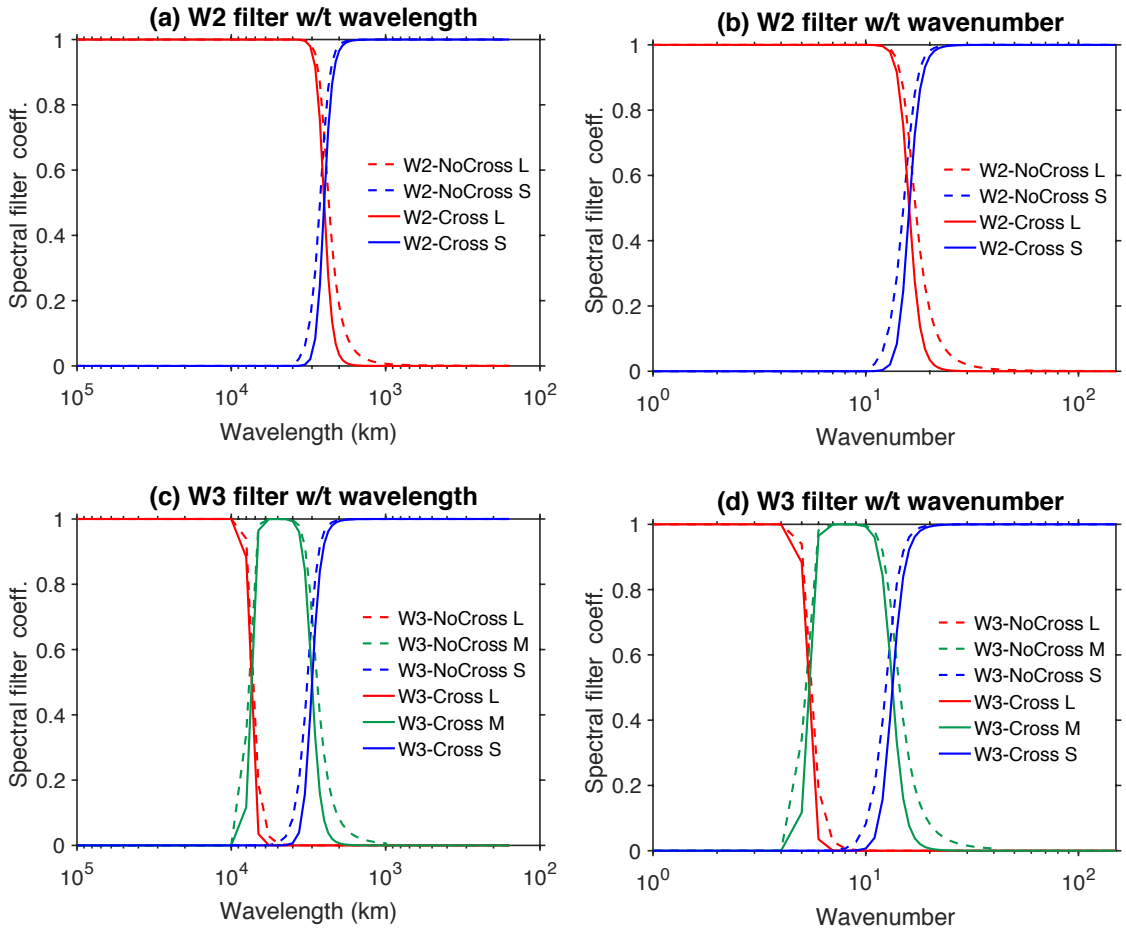

Figure 5.4 Spectral filter functions for SDL-NoCross (dashed) and SDL-Cross (solid) in the two- (top) and three- (bottom) waveband SDL experiments at large- (L, red), medium- (M, green) and small- (S, blue) scale wavebands as a function of the (a)(c) wavelength and (b)(d) total wavenumber. Note that the two-waveband SDL experiments in (a)(b) only contain the large- and small-scale wavebands.

To provide a robust comparison among different experiments, a paired permutation test was applied at 95% confidence level with 1000 replicates (Manly 2006). Procedures in Wang and Bishop (2005) were followed to collect independent samples for the significance test (see more details in Appendix B). For the global forecast comparison in sections 5.4.3, 5.4.4 and 5.5.1, the time series were first averaged in each of the independent subdomains distributed over the globe. This produced spatially independent time series. For the TC track forecast comparison

in sections 5.4.5 and 5.5.2, each storm was treated as an independent case. For the power spectral evaluation in section 5.6.3, each model level was treated as spatially independent. Then for each set of spatially independent time series, lag correlations were computed to further determine the length of a temporal block so that the temporal block series were weakly correlated in time. The objective of the aforementioned procedures is to determine effective degrees of freedom during the significance test. All sets of resultant temporally and spatially independent sample time series were pooled together to perform the paired permutation test. A false discovery rate (FDF) method (Wilks 2006) at 95% confidence level was further applied to ameliorate the simultaneous multiple hypothesis test issue.

5.4 Comparison of two-waveband SDL and scale-invariant localization experiments

5.4.1 Single observation experiment

To demonstrate the impacts of SDL, a single observation experiment was first performed for W1-1000, W1-300, W2-NoCross and W2-Cross (Fig. 5.5). A case featured with an interaction between a large-scale subtropical high and mesoscale typhoon was selected to more clearly reveal the differences between SDL-Cross and SDL-NoCross. W1-1000 applying much wider horizontal localization length scale produces two analysis increment maxima that are located at the observation location and to the north of the typhoon eye (Fig. 5.5a), respectively. The distant analysis increment maximum to the north of the typhoon eye suggests the raw ensemble covariances estimate an interaction between the subtropical high and typhoon. In addition, by applying the same horizontal localization length scale 1000 km *e*-folding distance for both large- and small-scale wavebands in W2-Cross, it reproduces the same analysis

increment pattern as W1-1000 (not shown here). This is consistent with the theory that applying the same amount of localization at different wavebands in SDL-Cross is equivalent to applying fixed uniform localization once at all scales (Buehner and Shlyaeva 2015). This further suggests that SDL was correctly implemented in the GSI-based 4DEnVar system. W1-300 applying much tighter horizontal localization length scale shows the most localized analysis increment pattern (Fig. 5.5b). Due to applying wider horizontal localization length scale for large-scale ensemble perturbations, W2-NoCross and W2-Cross show broader analysis increment patterns than W1-300. Compared to W1-1000, W2-NoCross and W2-Cross show overall more restricted analysis increment patterns due to much tighter horizontal localization applied for small-scale ensemble perturbations. W2-NoCross only shows one analysis increment maximum at the observation location, while W2-Cross that partially includes cross-waveband covariances maintains two analysis increment maxima as in W1-1000. This result suggests that the distant analysis increment maximum to the north of the typhoon eye in W1-1000 and W2-Cross is contributed by the cross-waveband covariances between the subtropical high and typhoon. The reduced magnitude of the analysis increment maxima in W2-Cross relative to W1-1000 can be attributed to the less-than-one peak in the cross-waveband localization matrix applied in SDL as shown in Fig. 5.1.

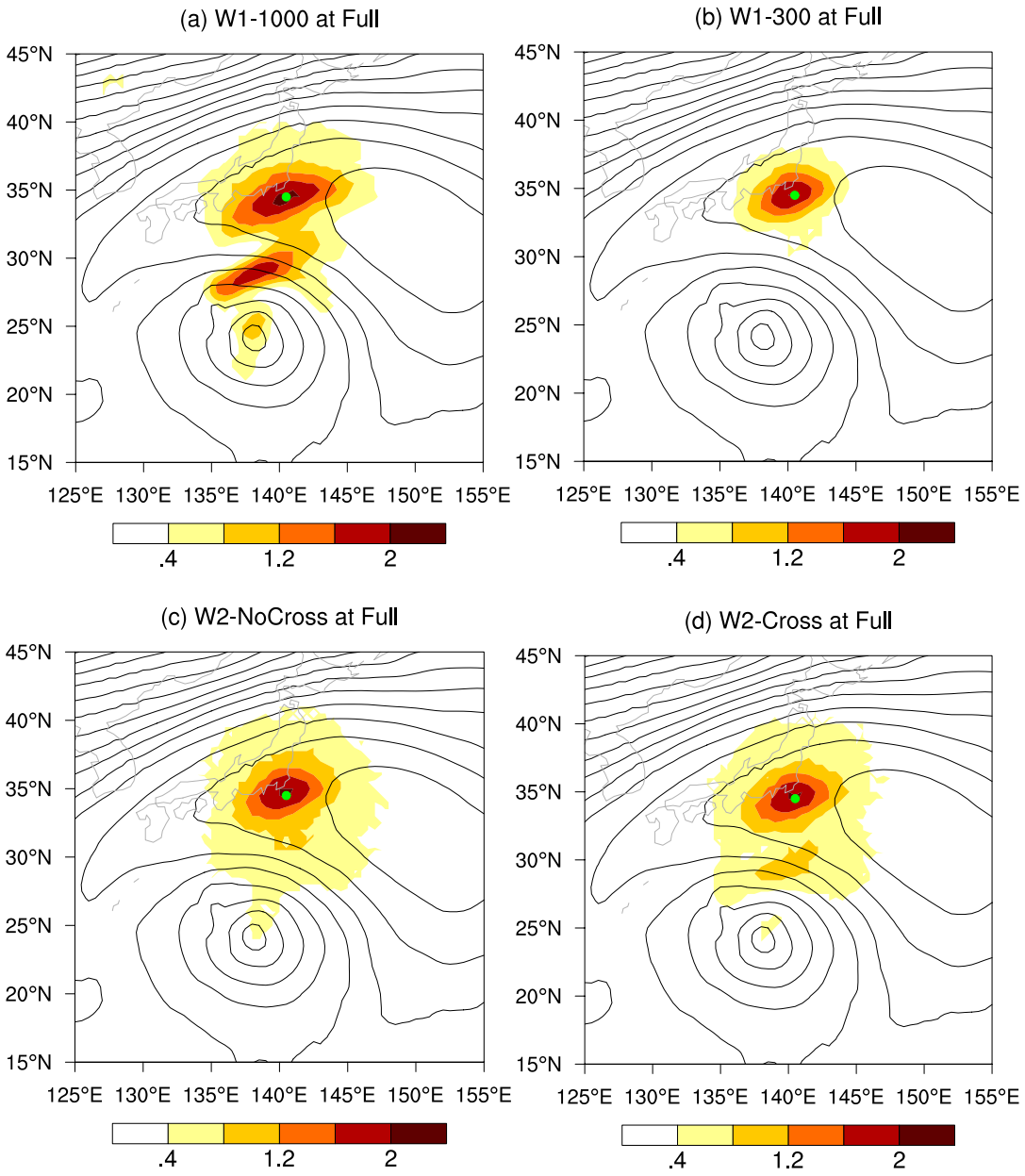

Figure 5.5 500 hPa zonal wind full-scale analysis increments in a hurricane example from assimilating a single zonal wind observation at the green dot that is 5 m s^{-1} higher than the background in (a) W1-1000, (b) W1-300, (c) W2-NoCross and (d) W2-Cross. The full-scale analysis increments in W2-NoCross and W2-Cross are combined analysis increments associated with decomposed large- and small-scale ensemble perturbations. The underneath grey contours denote the full-scale geopotential height at 500 hPa.

5.4.2 Analysis increment power

To investigate how the localization influences the analysis increments at different scales,

Fig. 5.6 shows the analysis increment power spectra as a function of total wavenumber for the temperature and wind variables at 500 hPa. By applying much wider horizontal localization length scale at 500 hPa (roughly at the 0.5 sigma level in Fig. 5.2), W1-1000 shows larger analysis increment power at all total wavenumbers compared to W1-300 and W1-Ope. W1-300 that applies slightly tighter horizontal localization at 500 hPa than W1-Ope shows slightly reduced analysis increment power. These results suggest that overall less tight horizontal localization generally produces larger analysis increment power in our cycled DA experiments. Furthermore, the analysis increment power in W2-NoCross and W2-Cross is at a magnitude closer to W1-1000 (W1-300) at small (large) total wavenumbers. This shall be expected because the two-waveband SDL experiments apply the same horizontal localization length scale as W1-1000 (W1-300) at small (large) total wavenumbers. The analysis increment power differences between W2-NoCross and W2-Cross are slightly more noticeable at small total wavenumbers.

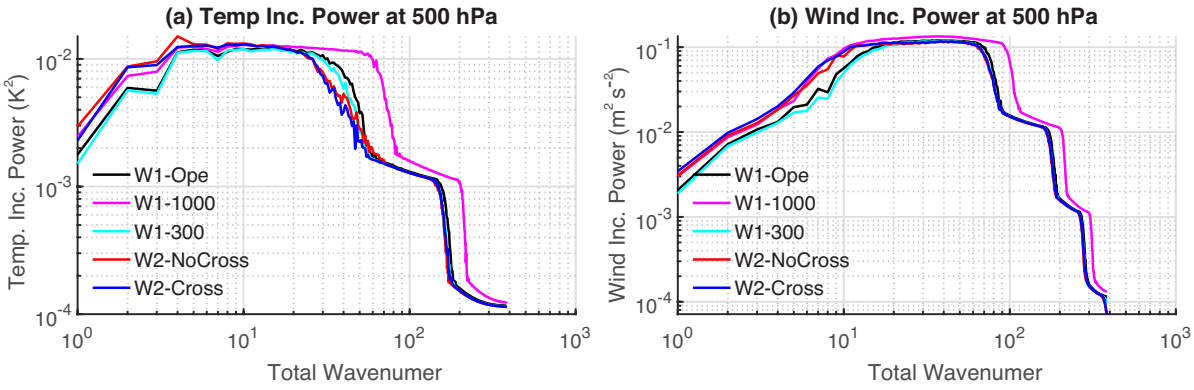

Figure 5.6 Analysis increment power spectrum as a function of total wavenumber for the (a) temperature (unit: K^2) and (b) wind (unit: $m^2\ s^{-2}$) variables at 500 hPa in W1-Ope (black), W1-1000 (magenta), W1-300 (cyan), W2-NoCross (red) and W2-Cross (blue).

5.4.3 Forecast verification against rawinsonde observations

Root-mean-square-errors (RMSEs) were calculated by comparing the 6-hour background forecasts against the rawinsonde observations. In both sections 5.4.3 and 5.4.4, W1-Ope is used as a reference for comparison. As discussed in section 5.3, W1-Ope adopts the operationally-tuned, level-dependent, scale-invariant localization, and therefore provides a strict reference. By applying much wider horizontal localization length scale for the full-scale ensemble perturbations, W1-1000 shows the worst 6-hour temperature and wind forecasts at most model levels. Compared to W1-Ope, W1-300 shows comparable or slightly improved 6-hour background forecasts at several model levels below 100 hPa especially for the wind forecasts. It suggests that slightly reducing horizontal localization length scale below 100 hPa in W1-Ope may further benefit the forecasts at such model levels in the operational hybrid 4DEnVar system. However, W1-300 produces larger temperature forecast error than W1-Ope above 100 hPa. It is also noticed that W1-1000 has the least degradation relative to W1-Ope above 100 hPa compared to lower model levels. These results together suggest that a wider horizontal localization length scale for the full-scale ensemble perturbations is generally beneficial at upper model levels. This is consistent with the current operational level-dependent horizontal localization settings in W1-Ope, and may explain the degradation in W1-300 above 100 hPa that applies much tighter horizontal localization.

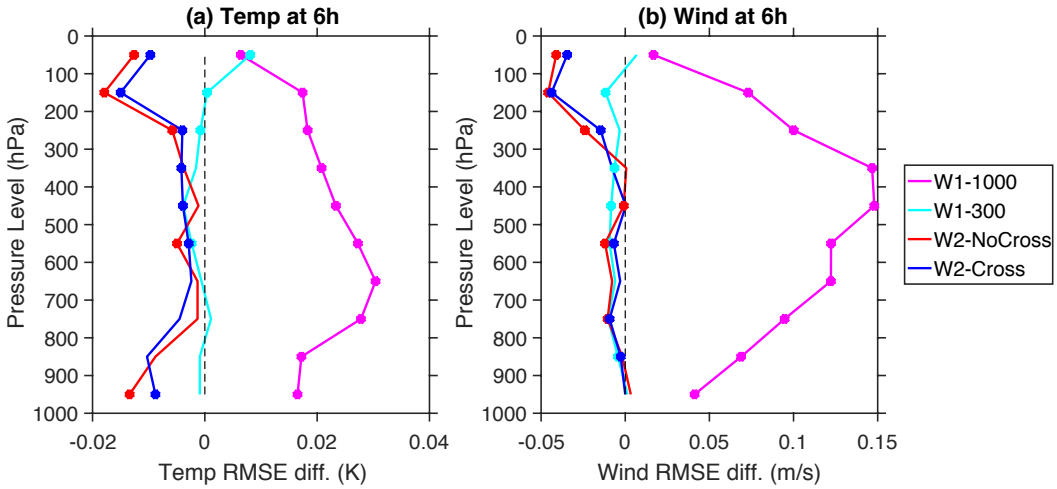

Figure 5.7 Difference of the Root-mean-square-error (RMSE) of the 6-hour (a) temperature (Unit: K) and (b) wind background (Unit: m s^{-1}) forecasts against the rawinsondes as a function of pressure level in W1-1000 (magenta), W1-300 (cyan), W2-NoCross (red) and W2-Cross (blue) relative to W1-Ope. Negative (positive) values mean improved (degraded) 6-hour background forecasts relative to W1-Ope. The bold dot sign indicates the RMSE in a particular experiment is significantly different from W1-Ope at or above 95% confidence level by applying a paired permutation test with 1000 replicates combined with the FDR method at 95% confidence level. The dashed black line denotes zero values.

On the other hand, W2-NoCross and W2-Cross generally improve the 6-hour global forecasts over W1-Ope, W1-1000 and W1-300 at most model levels, suggesting the benefits of SDL. In particular, W2-NoCross and W2-Cross that apply tighter horizontal localization for the small-scale ensemble perturbations show the largest improvement for 6-hour temperature and wind forecasts over W1-Ope above 200 hPa. This can be further related to that W1-Ope starts to increase horizontal localization length scale above 200 hPa. Therefore, wider horizontal localization length scale for the large-scale ensemble perturbations in the two-waveband SDL experiments may contribute more above 200 hPa. This is also consistent with our earlier discussion that wider horizontal localization length scale is beneficial at upper model levels. Relative performances between W2-NoCross and W2-Cross are a bit mixed for the 6-hour

background forecasts below 300 hPa. However, W2-NoCross slightly improves the 6-hour background forecasts more than W2-Cross above 300 hPa, suggesting the cross-waveband covariances estimated by the ensemble may not be reliable above 300 hPa. This is consistent with Caron et al. (2019) that relative performances between SDL-NoCross and SDL-Cross depend on the accuracy of estimated cross-waveband covariances.

5.4.4 Global forecast verification against ECMWF reanalysis

To evaluate medium-range global forecasts, root-mean-square-errors (RMSEs) between the five-day global forecasts and ECMWF reanalyses were calculated at selected model levels every 6 hours. Figure 5.8 shows the RMSE difference relative to W1-Ope as a function of forecast lead time and model pressure level. W1-1000 in general shows the worse five-day global temperature and wind forecasts than W1-Ope. This further suggests negative impacts of applying much wider horizontal localization length scale at these selected model levels in W1-1000 than W1-Ope. Different from the consistently degraded global wind forecasts, W1-1000 shows slightly better temperature forecasts than W1-Ope at several lower model levels within two days and at 10 hPa beyond two days. This may suggest the necessity of considering variable-dependency of the localization. Compared to W1-Ope, W1-300 degrades global temperature forecasts above 150 hPa. This is hypothesized to result from the negative impacts of applying too tight horizontal localization at such levels in W1-300 relative to W1-Ope. Compared to the global wind forecasts, the more degraded global temperature forecasts above 150 hPa in W1-300 may also suggest that the optimal horizontal localization length scale is variable-dependent. Below 150 hPa, W1-300 produces slightly more accurate or comparable temperature forecasts

relative to W1-Ope. However, W1-300 shows more accurate global wind forecasts up to 50 hPa at shorter forecast lead times. This is consistent with the 6-hour background forecast verification (Fig. 5.7). It again suggests the potential additional benefits of slightly reducing horizontal localization length scale at lower model levels in the operational hybrid 4DEnVar system for the global wind forecasts. Different from W1-1000 and W1-300, W2-NoCross and W2-Cross that apply SDL consistently improve the global forecasts over W1-Ope at most model levels and forecast lead times. It is noticed that W2-NoCross shows slightly degraded temperature forecasts relative to W1-Ope below 850 hPa between three and four days. This may suggest that the current scale separation based on the wind variable may not be optimal for the temperature variable especially when applying SDL-NoCross.

Figures 5.9a,b further compare the five-day global forecasts between W2-NoCross and W2-Cross. W2-NoCross tends to show slightly better global forecasts than W2-Cross within one-day forecast lead times. Beyond one day, however, W2-Cross improves the global forecasts over W2-NoCross. The advantage of W2-Cross at shorter forecast lead times may be associated with the local spatial averaging of ensemble covariances in SDL-NoCross, due to its complete removal of cross-waveband covariances. While the local spatial averaging in SDL-NoCross may help alleviate sampling error and improve fitting of the analysis to the observations, it would likely retain less heterogeneity of ensemble covariances (Caron et al. 2019). In contrast, SDL-Cross partially includes the cross-waveband covariances and retains more heterogeneity of ensemble covariances. The retained more heterogeneity of ensemble covariances in W2-Cross, together with its more balanced analysis (in Fig. 5.16 in section 5.6.2) may explain its more accurate global forecasts at longer forecast times than W2-NoCross.

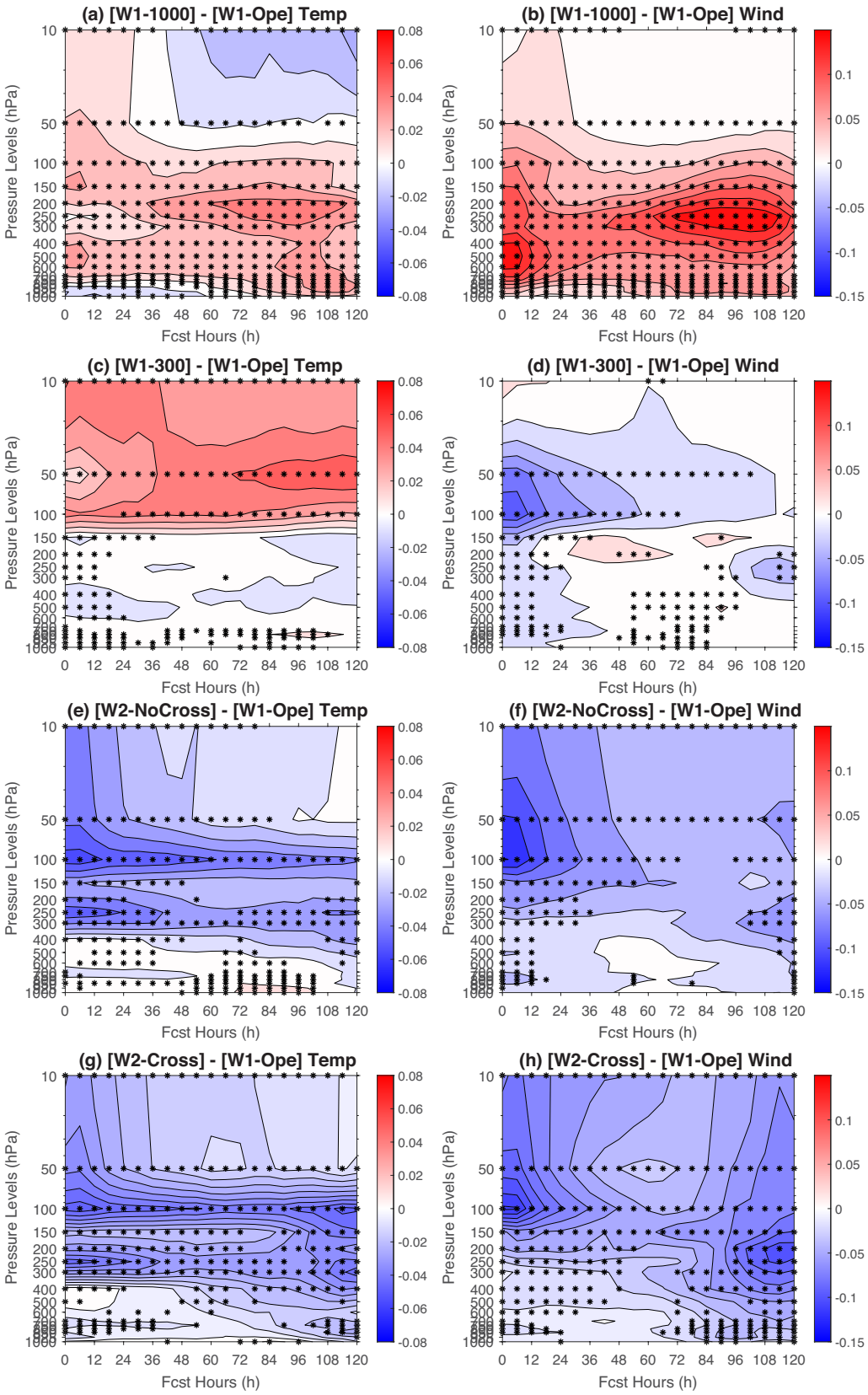
(a) [W1-1000] - [W1-Ope] Temp
(b) [W1-1000] - [W1-Ope] Wind
(c) [W1-300] - [W1-Ope] Temp
(d) [W1-300] - [W1-Ope] Wind
(e) [W2-NoCross] - [W1-Ope] Temp
(f) [W2-NoCross] - [W1-Ope] Wind
(g) [W2-Cross] - [W1-Ope] Temp
(h) [W2-Cross] - [W1-Ope] Wind
Pressure Levels (hPa)
Fcst Hours (h)

Figure 5.8 Globally and temporally averaged root-mean-square-error (RMSE) difference from W1-Ope for the (left) temperature (unit: K) and (right) wind (unit: $m\,s^{-1}$) forecasts in (a)(b) W1-1000, (c)(d) W1-300, (e)(f) W2-NoCross and (g)(h) W2-Cross that were verified against ECMWF reanalyses as a function of forecast time to 5 days on the horizontal axis and pressure level on the vertical axis. Blue (red) color indicates the improved (degraded) forecasts relative to W1-Ope. The asterisk signs at the corresponding forecast times and pressure levels indicate that the RMSE difference from W1-Ope is statistically significant at or above 95% confidence level by applying the paired permutation test with 1000 replicates combined with the FDR method at 95% confidence level.

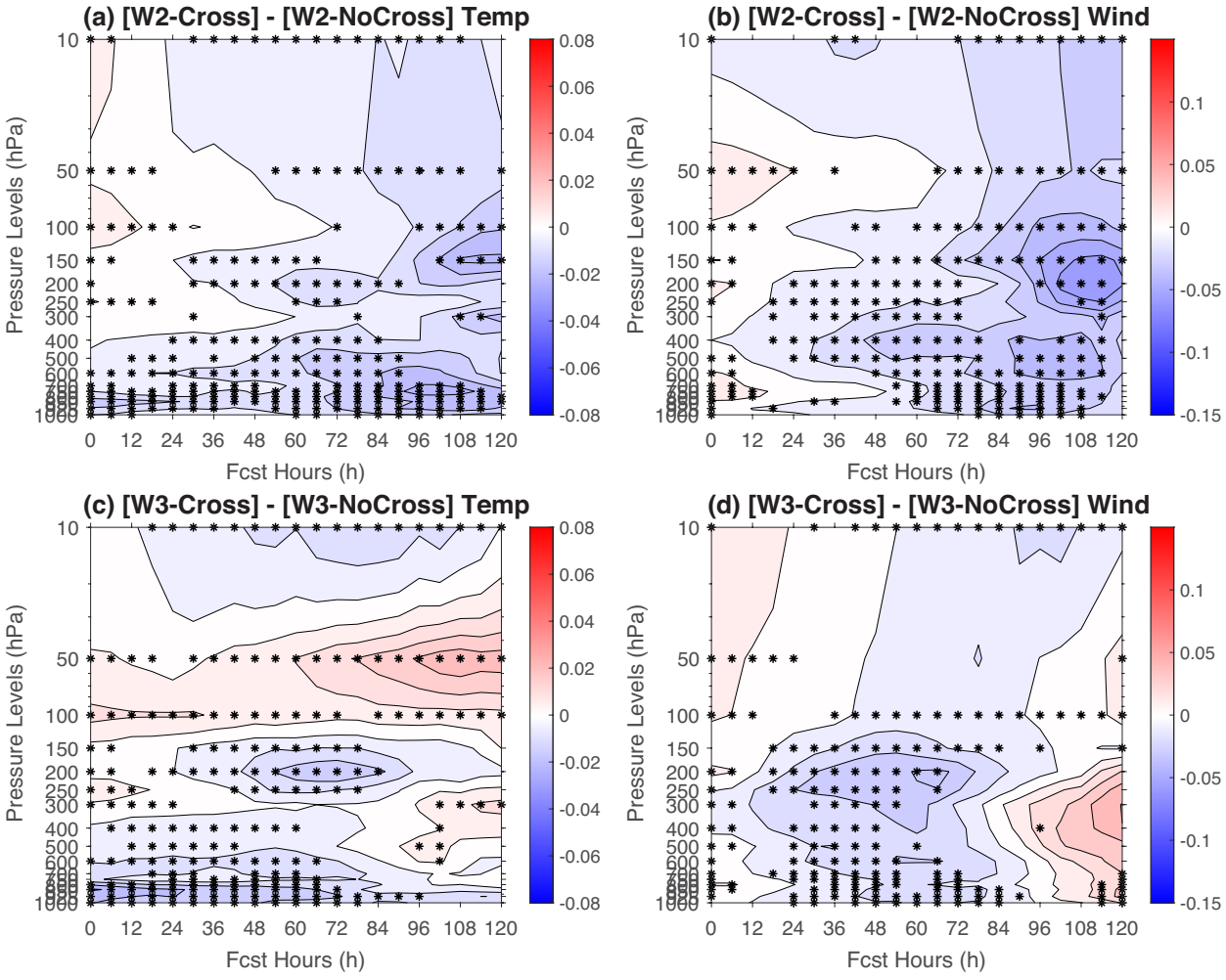

Figure 5.9 As in Fig. 5.8 but for the global (left) temperature and (right) wind forecast RMSE difference between the experiments applying SDL-Cross and SDL-NoCross in the (a)(b) two- and (c)(d) three-waveband SDL experiments. Blue (red) color indicates SDL-Cross shows improved (degraded) forecasts compared to SDL-NoCross.

5.4.5 TC track forecast verification

As discussed in the introduction, the prediction of the tropical cyclone (TC) track requires the TC, its embedded large-scale environment and their interaction to be analyzed properly. Therefore, in this study, we investigate how SDL influences the TC track forecasts.

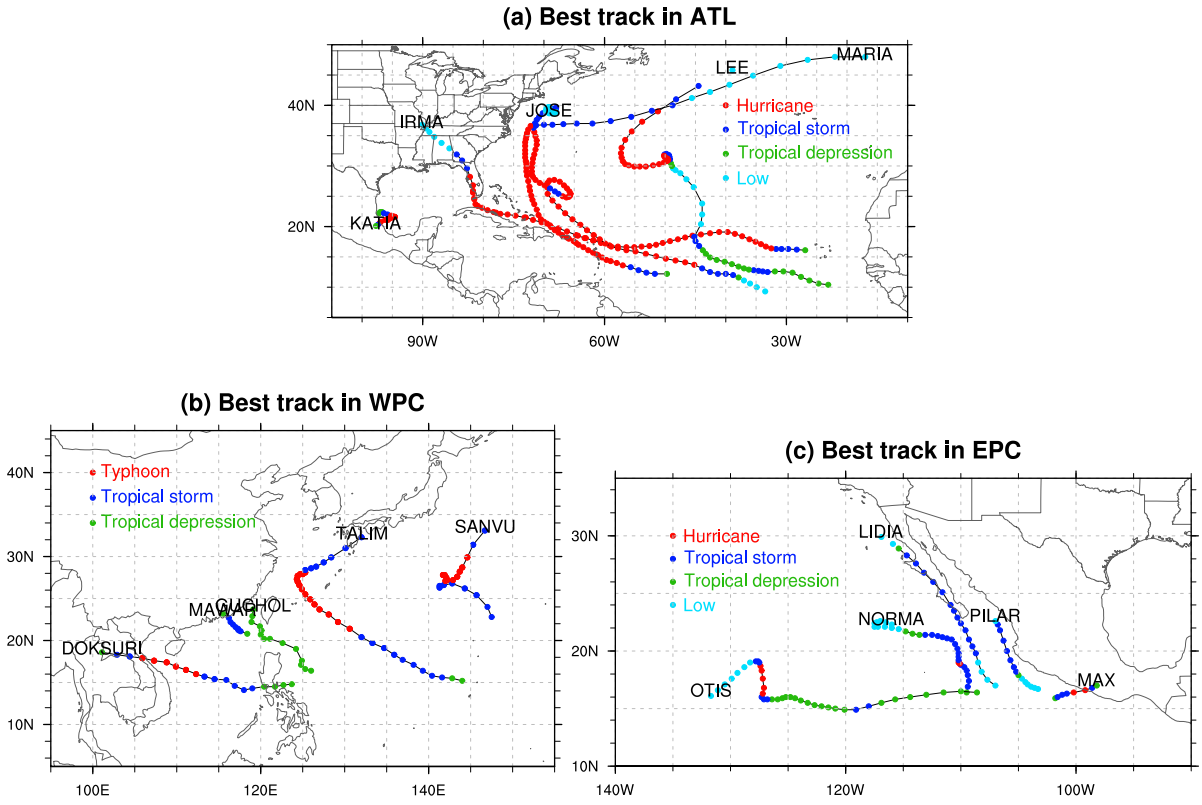

Figure 5.10 Best track of the tropical cyclones during the experiment period in the (a) Atlantic, (b) Western Pacific and (c) Eastern Pacific basins.

In the cycled DA experiment period, there were fifteen named TCs in the Northern Atlantic basin and Northern Pacific basin. Ten TCs reached the typhoon or hurricane category (Fig. 5.10). The NCEP TC tracker (Marchok 2002) was used to calculate the forecasted TC location. Following Wang and Lei (2014) and Huang and Wang (2018), the TC track error was calculated against the best track data (National Hurricane Center and Central pacific Hurricane

Center 2020; Naval Oceanography Portal 2020), and averaged over all fifteen TCs. Figure 5.11a shows the TC track error difference relative to W2-Cross as a function of forecast lead time. W1-Ope improves the TC track forecasts over W1-1000 almost out to five days, and over W1-300 beyond three days. On the other hand, W2-Cross and W2-NoCross show significantly improved TC track forecasts over W1-1000 out to four days, and over W1-300 beyond three days. These results suggest improved analyses in W2-Cross and W2-NoCross relative to scale-invariant localization experiments lead to improved subsequent TC track forecasts. Furthermore, W2-Cross improves the TC track forecasts over W2-NoCross beyond three days. This is consistent with better global forecasts in W2-Cross relative to W2-NoCross at longer forecast lead times in Figs. 5.9a,b. However, the TC track forecasts between W1-Ope and W2-Cross are statistically indistinguishable. This is not the case in their global forecast comparison in Figs. 5.8g,h. It is hypothesized that the current scale separation in the two-waveband SDL experiments may not be sufficient to improve the TC track forecasts. Impact of SDL on TC track forecasts will be further examined in the three-waveband SDL experiments in section 5.5.2. Percentages of the track forecasts that are more accurate than W2-Cross (Wang and Lei 2014) in Fig. 5.11b are consistent with the TC track error differences in Fig. 5.11a. For instance, at the forecast lead times beyond two and a half days, more than 50% of the track forecasts in W2-Cross are more accurate than W1-1000, W1-300 and W2-NoCross.

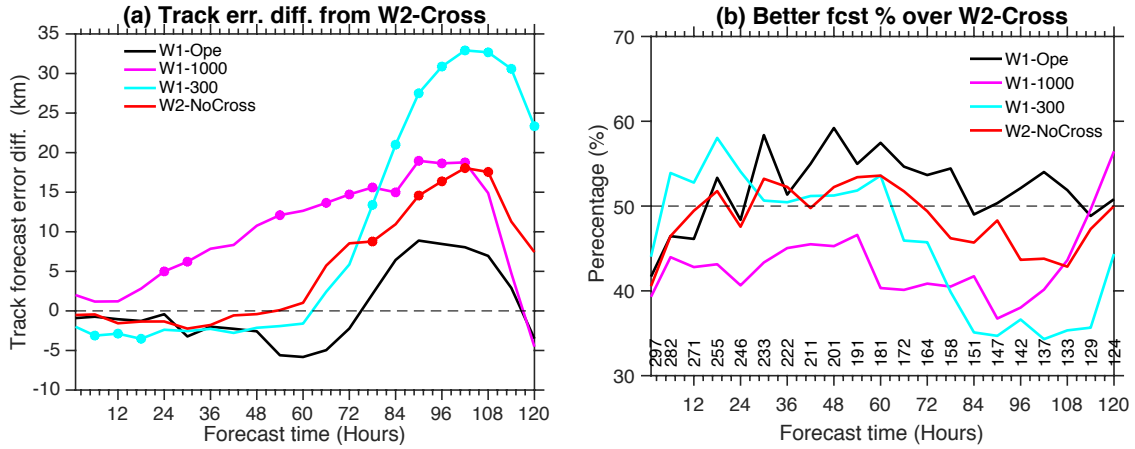

Figure 5.11 Figure 11 (a) Track forecast error difference in W1-Ope (black), W1-1000 (magenta), W1-300 (cyan), W2-NoCross (red) relative to W2-Cross. The bold dots on the curves in (a) indicate that the track error difference from W2-Cross is statistically significant at or above 95% confidence level by applying the paired permutation test with 1000 replicates combined with the FDR method at 95% confidence level at the corresponding forecast lead time. (b) Percentage of the track forecasts that are more accurate than W2-Cross with the same line styles and color indexes in (a). The numbers right above the horizontal axis in (b) denote the sample size at the corresponding forecast lead time.

5.5 Comparison of two- and three-waveband SDL experiments

To investigate how SDL performs in response to the number of decomposed wavebands, the two- and three-waveband SDL experiments were further compared in this subsection. As discussed in the introduction, the major difference in the two- and three-waveband SDL experiment designs is tighter horizontal localization applied at medium-scale waveband. Consistent with the two-waveband SDL experiments, the three-waveband SDL experiments improve the global temperature and wind forecasts over W1-Ope at most model levels and forecast lead times (not shown here). This section will focus on the comparison among the two- and three-waveband SDL experiments.

5.5.1 Global forecast verification against ECMWF reanalysis

Figure 5.12 shows the RMSE difference of the global temperature and wind forecasts against the ECMWF analyses between the two- and three-waveband SDL experiment counterparts. Above 50 hPa, the three-waveband SDL experiments in general produce less accurate global forecasts over five days than their two-waveband SDL experiment counterparts. As discussed earlier, wider overall horizontal localization length scale is beneficial at upper model levels. Reduced overall horizontal localization length scale in the three-waveband SDL experiments, due to tighter horizontal localization applied at medium-scale waveband, may be the reason for their degraded global forecasts at upper model levels. On the other hand, overall tighter horizontal localization length scale is beneficial at lower model levels. So reduced overall horizontal localization length scale in the three-waveband SDL experiments may explain their better global temperature and wind forecasts below 100 hPa at least out to 3 days compared to the two-waveband SDL experiment counterparts. In particular, the advantage of W3-NoCrsoss relative to W2-NoCross below 100 hPa almost lasts for the entire five days, while that of W3-Crsoss relative to W2-Cross only lasts out to three days. As shown later in Fig. 5.16, the analysis in the three-waveband SDL experiments is more imbalanced than the two-waveband SDL experiment counterparts. At longer forecast lead times, the negative impacts from less balanced analysis may overrule the positive benefits of more accurate analysis resultant from overall tighter horizontal localization. Furthermore, the amount of imbalance increased going from two wavebands to three wavebands is larger in the experiments applying SDL-Cross than those applying SDL-NoCross. This likely explains why the benefit of W3-Cross versus W2-Cross below 100 hPa only lasts up to 3 days.

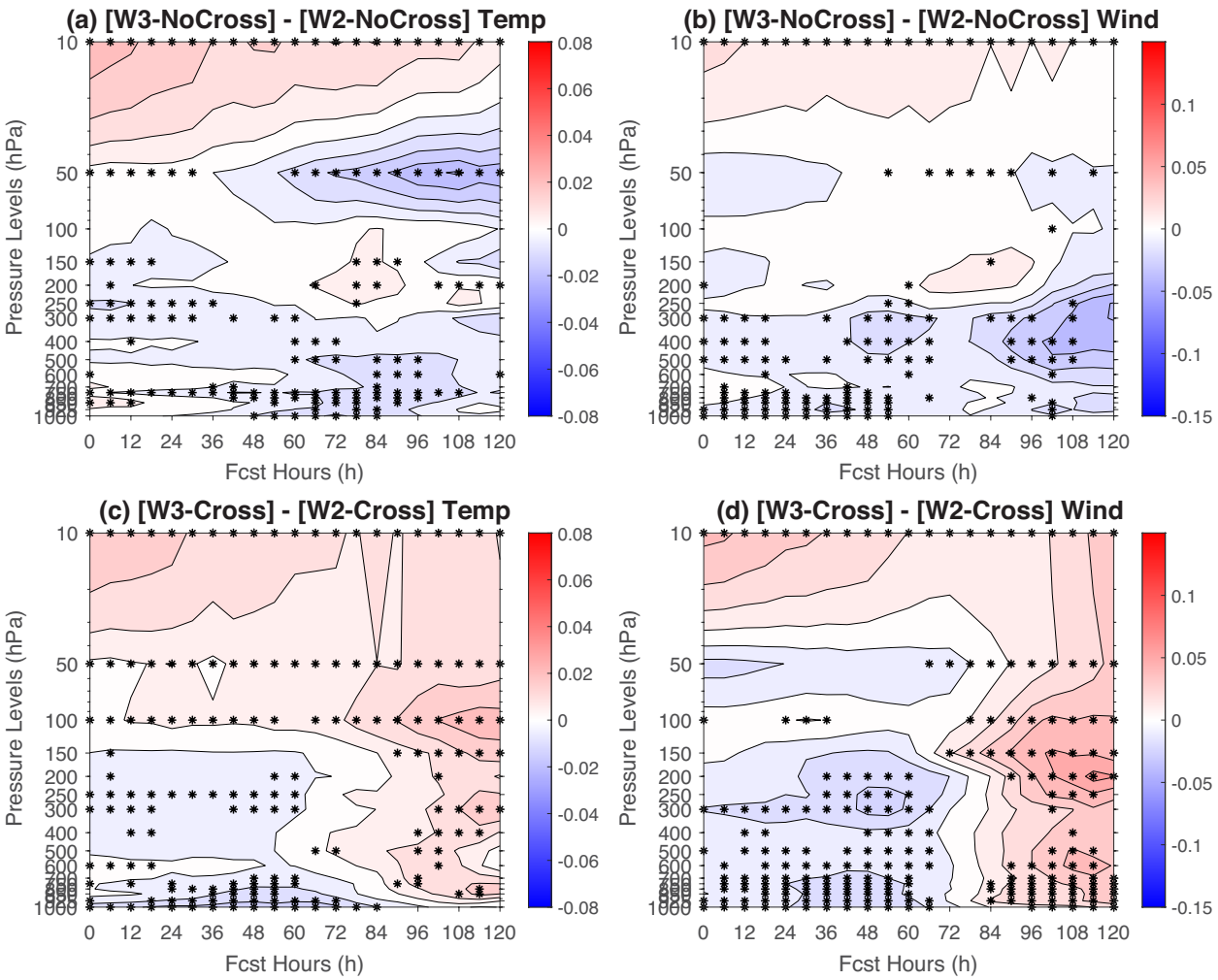

Figure 5.12 As in Fig. 5.8 but for the global (left) temperature and (right) wind forecast RMSE difference between the two- and three-waveband SDL experiment counterparts that apply (a)(b) SDL-NoCross and (c)(d) SDL-Cross. Blue (red) color indicates the three-waveband SDL experiment shows improved (degraded) forecasts compared to its two-waveband experiment counterpart.

Figures 5.9c,d show the forecast RMSE difference between W3-NoCross and W3-Cross. As in the two-waveband SDL experiments, W3-NoCross produces slightly more accurate global wind forecasts within twelve hours than W3-Cross. As the forecast lead time increases, the

global wind forecasts in W3-Cross becomes more accurate than W3-NoCross. This variation of the wind forecast performances of W3-NoCross relative to W3-Cross as the forecast lead time increases also happens to the global temperature forecasts at model levels between 400 hPa and 150 hPa, and at about 10 hPa. As discussed in the two-waveband SDL experiments in section 5.4.4, the better global forecasts of W3-Cross over W3-NoCross at longer forecast lead times may be associated with its retained more heterogeneity of ensemble covariances and more balanced analysis (Fig. 5.16). However, it is also noted that the outperformance of W3-Cross over W3-NoCross lasts for a shorter period of time than that of W2-Cross over W2-NoCross. This may be related to less amount of imbalance reduction between SDL-Cross and SDL-NoCross when increasing from two wavebands to three wavebands. Between 100 hPa and 50 hPa, W3-Cross shows worse global temperature forecasts than W3-NoCross over five days. However, this is neither the case between W2-NoCross and W2-Cross, nor the case for the wind forecasts. These results suggest that relative performances between SDL-NoCross and SDL-Cross, owing to the partial cross-waveband covariances in SDL-Cross, could vary with the number of decomposed wavebands in SDL, model level, and model variable.

5.5.2 TC track forecast verification

Figure 5.13 shows the TC track forecast error difference in the two- and three-waveband SDL experiments relative to W2-Cross, and the percentages of the TC track forecasts that are better than W2-Cross. W1-Ope is included in Fig. 5.13 for further comparison with the three-waveband SDL experiments for the TC track forecasts. W3-Cross significantly improves the TC track forecasts over W1-Ope and W2-Cross within two-day forecast lead times. W3-NoCross in

general shows more accurate TC track forecasts than W2-NoCross between one-and-half and three days. Improved TC track forecasts in the three-waveband SDL experiments over their two-waveband SDL counterparts may benefit from the three-waveband scale separation and tighter horizontal localization length scale applied at medium-scale waveband. For example, the decomposed wind background displays more representative and distinguished features in the three-waveband SDL experiments (Fig. 5.3). However, W3-Cross shows statistically less accurate TC track forecast than W2-Cross between three and four days. This is consistent with its general less accurate global forecasts in W3-Cross than W2-Cross at such forecast time periods. Furthermore, the percentage differences in Fig. 5.13b are generally consistent with the TC track error differences in Fig. 5.13a. For instance, more than 50% percent of track forecasts in W3-Cross are more accurate than W2-Cross between twelve hours and two and half days.

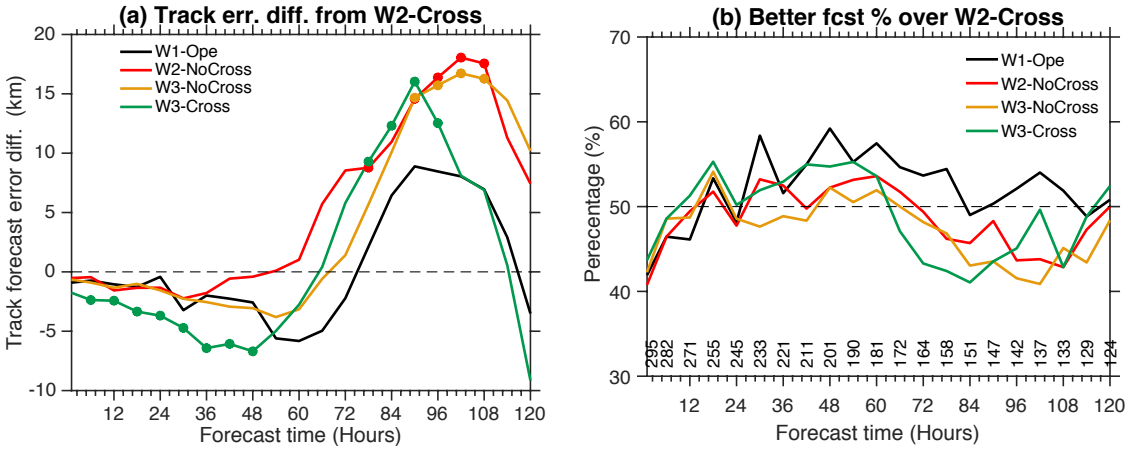

Figure 5.13 As in Fig. 5.11 but for (a) the track forecast error difference and (b) percentage of more accurate track forecasts in W1-Ope (black), W2-NoCross (red), W3-NoCross (orange), W3-Cross (green) in contrast to W2-Cross.

5.6 Additional diagnostics to understand performance differences

5.6.1 Localized correlation pattern comparison

To have a systematic evaluation of localized correlation patterns from all the localization methods, a total of 200 localized zonal wind correlation samples at 500 hPa were collected over the globe for all the seven experiments. Figure 5.14 shows 45 localized correlation samples. Given slightly broader horizontal localization length scale applied in W1-Ope at 500 hPa relative to W1-300 (Fig. 5.2), its localized correlation patterns are only slightly broader than those in W1-300 and are thus not included in Fig. 5.14. In general, the localized correlation patterns vary with latitudes, geographical locations and weather systems (Buehner 2012b). W1-1000 applying much wider horizontal localization length scale shows the most broad localized correlation patterns, and retains the largest amount of noises from sampling error and largest amount of heterogeneity. As expected, W1-300 shows the tightest localized correlation patterns. Due to tighter horizontal localization length scale applied at small- and/or medium-scale waveband, the localized correlation patterns in the SDL experiments are tighter, smoother and retain less heterogeneity of ensemble correlations compared to W1-1000. On the other hand, the SDL experiments show broader localized correlation patterns than W1-300 due to their wider horizontal localization length scale applied at large- and/or medium-scale waveband. Owing to the local spatial averaging of ensemble covariances, the localized correlation patterns from SDL-NoCross tend to be less heterogenetic than those from SDL-Cross. By applying tighter horizontal localization length scale at medium-scale waveband, the three-waveband SDL experiments produce more localized correlation patterns than their two-waveband SDL counterparts.

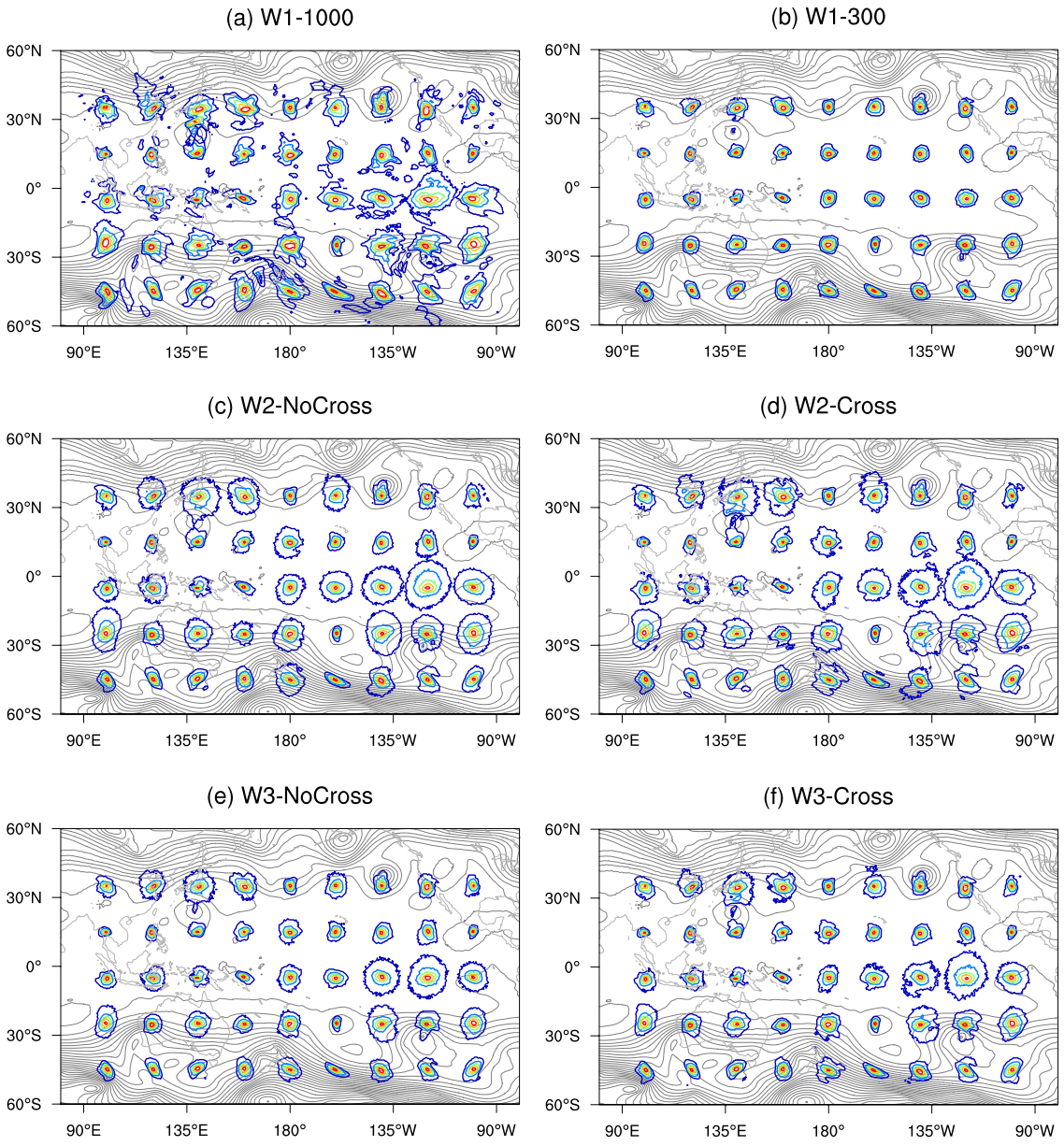

Figure 5.14 45 localized zonal wind correlation samples at 500 hPa distributed over the globe in (a) W1-1000, (b) W1-300, (c) W2-NoCross, (d) W2-Cross, (e) W3-NoCross and (f) W3-Cross. The thin grey contours give the geopotential height at 500 hPa. The colored contours display the correlation magnitude from 0.15 to 0.95 with an interval of 0.2.

To further quantify and compare the spatial variability of localized ensemble correlations, the standard deviation of 200 correlation samples was calculated as a function of distance (Fig.

5.15). In the experiments applying scale-invariant localization, W1-1000 (W1-300) shows the largest (least) standard deviation. It suggests that tighter localization retains less spatial variation of raw ensemble correlations. By applying tighter horizontal localization at small- and/or medium-scale waveband, all the SDL experiments have smaller standard deviation than W1-1000. Due to applying broader horizontal localization length scale at large- and/or medium-scale waveband, the SDL experiments show larger standard deviation than W1-300 beyond 300 km. However, within 300 km, W2-Cross and W3-Cross show comparable standard deviation as W1-300, while W2-NoCross and W3-NoCross have smaller standard deviation than W1-300. More noticeably, the experiments applying SDL-Cross that partially includes cross-waveband correlations consistently show larger standard deviation than those applying SDL-NoCross. Finally, the three-waveband SDL experiments show less standard deviation or variability than their two-waveband SDL experiment counterparts at distance between 300 km and 1,700 km.

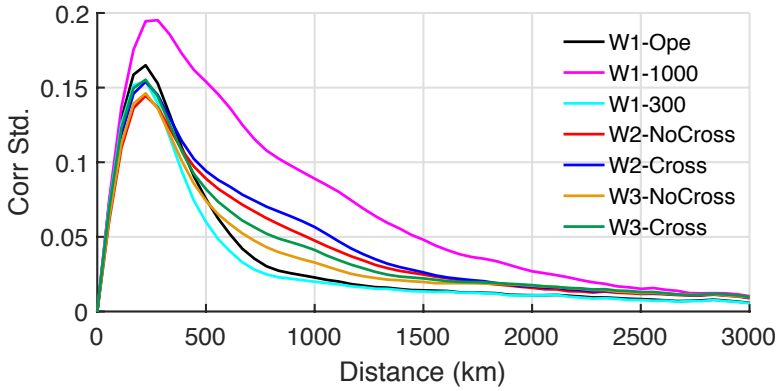

Figure 5.15 Standard deviation of 200 samples of localized zonal wind correlations as a function of distance on the horizontal axis that were collected from W1-Ope (black), W1-1000 (magenta), W1-300 (cyan), W2-NoCross (red), W2-Cross (blue), W3-NoCross (orange) and W3-Cross (green).

5.6.2 Impact of dynamical balance

Covariance localization causes imbalance in the resultant analysis in the ensemble-based DA system (Wang et al. 2007; Holland and Wang 2013; Wang et al. 2013; Lei and Whitaker 2016; Caron et al. 2019). Direct impact from the imbalance on the subsequent forecasts is expected to decrease as the forecast time increases. However, the errors caused by the imbalance can grow with time and indirectly influence the forecast accuracy at longer forecast lead times (Wang et al. 2013; Wang and Lei 2014; Lei and Whitaker 2016). To measure the imbalance, the absolute hourly pressure tendency (Lynch and Huang 1992) was calculated for all the seven experiments (Fig. 5.16). W1-Ope applying the operationally-tuned, level-dependent, scale-invariant horizontal localization is the most balanced among all the seven experiments. W1-300 applying tighter horizontal localization length scale is less balanced than W1-Ope. This is within the general expectation that stronger localization introduces more imbalance. However, W1-1000 is less balanced than W1-Ope and W1-300. This seemingly out-of-expectation result is consistent with Greybush et al. (2011) who hypothesized that larger analysis increments that are required to correct inaccurate background have the potential of producing more imbalance through the DA update. As shown in Figs. 5.6 and 5.7, W1-1000 generally shows the worst background forecasts, but the largest analysis increments compared to W1-Ope and W1-300.

The analyses in the SDL experiments are less balanced than W1-Ope, which could be because the overall effective horizontal localization length scale in the SDL experiments is smaller especially at upper model levels. Interestingly, the analysis in SDL-Cross is more balanced than SDL-NoCross, likely due to the partial maintenance of the cross-waveband covariances in SDL-Cross. The more balanced analysis in SDL-Cross may contribute to its

general better forecasts at longer forecast lead times than SDL-NoCross in section 5.5. Furthermore, the three-waveband SDL experiments are less balanced than their two-waveband SDL experiment counterparts. It may result from the reduced overall effective horizontal localization length scale in the three-waveband SDL experiments.

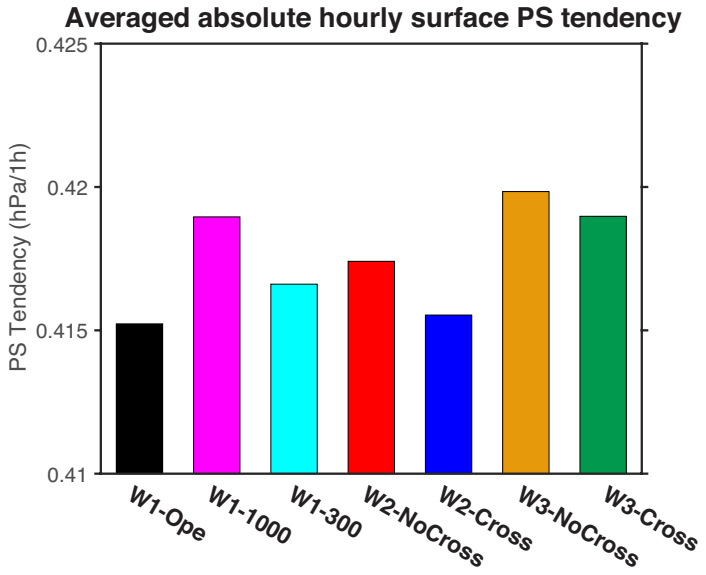

Figure 5.16 Globally averaged absolute hourly surface pressure tendency (hPa $1h^{-1}$) in W1-Ope (black), W1-1000 (magenta), W1-300 (cyan), W2-NoCross (red), W2-Cross (blue), W3-NoCross (orange) and W3-Cross (green).

5.6.3 Forecast error comparison as a function of total wavenumber

To investigate how SDL affects forecast error at different scales, the errors of the global forecasts against the ECMWF reanalyses in Fig. 5.8 were decomposed in spectral space and displayed as a function of total wavenumber. A representative total energy norm was performed on decomposed errors (Wang and Bishop 2003), and averaged over all selected model levels in Fig. 5.8. Figure 5.17 shows the error total energy difference relative to W1-Ope as a function of forecast lead time and total wavenumber. By applying wider horizontal localization length scale below 100 hPa, W1-1000 (Fig. 5.17a) shows larger error total energy than W1-Ope at most total

wavenumbers and forecast lead times. The smaller error total energy of W1-1000 versus W1-Ope at total wavenumbers below ten within two days may be because wider horizontal localization length scale contributes positively to the large-scale component of ensemble covariances and resultant analysis. In comparison, since tighter localization is beneficial for the small-scale component of ensemble covariances and resultant analysis, W1-300 (Fig. 5.17b) generally shows smaller error total energy than W1-Ope at total wavenumbers above ten. At total wavenumbers below ten, W1-300 shows comparable or slightly larger error total energy than W1-Ope.

In addition, the two- and three-waveband SDL experiments show smaller error total energy than W1-Ope at most total wavenumbers and forecast lead times especially for the experiments applying SDL-Cross. Their largest error total energy reduction relative to W1-Ope appears at total wavenumbers between five and twenty beyond three days where W1-Ope shows the maximum error total energy (not shown here). Without considering cross-waveband covariances, the experiments applying SDL-NoCross tend to have slightly larger error total energy than those applying SDL-Cross at longer forecast lead times. For example, the experiments applying SDL-NoCross even show slightly larger error total energy than W1-Ope at total wavenumbers of three and four beyond two days. But this is not the case for the experiments applying SDL-Cross, suggesting additional benefits of partially including cross-waveband covariances. By applying tighter horizontal localization at medium-scale waveband, the three-waveband SDL experiments generally show slightly larger error total energy than their two-waveband SDL counterparts especially at total wavenumbers between five and twenty that correspond to the error total energy maximum in W1-Ope. Overall, the SDL experiments

especially applying SDL-Cross show the potential to improve the global forecasts over W1-Ope almost at all total wavenumbers in contrast to a subset of total wavenumbers in W1-1000 and W1-300.

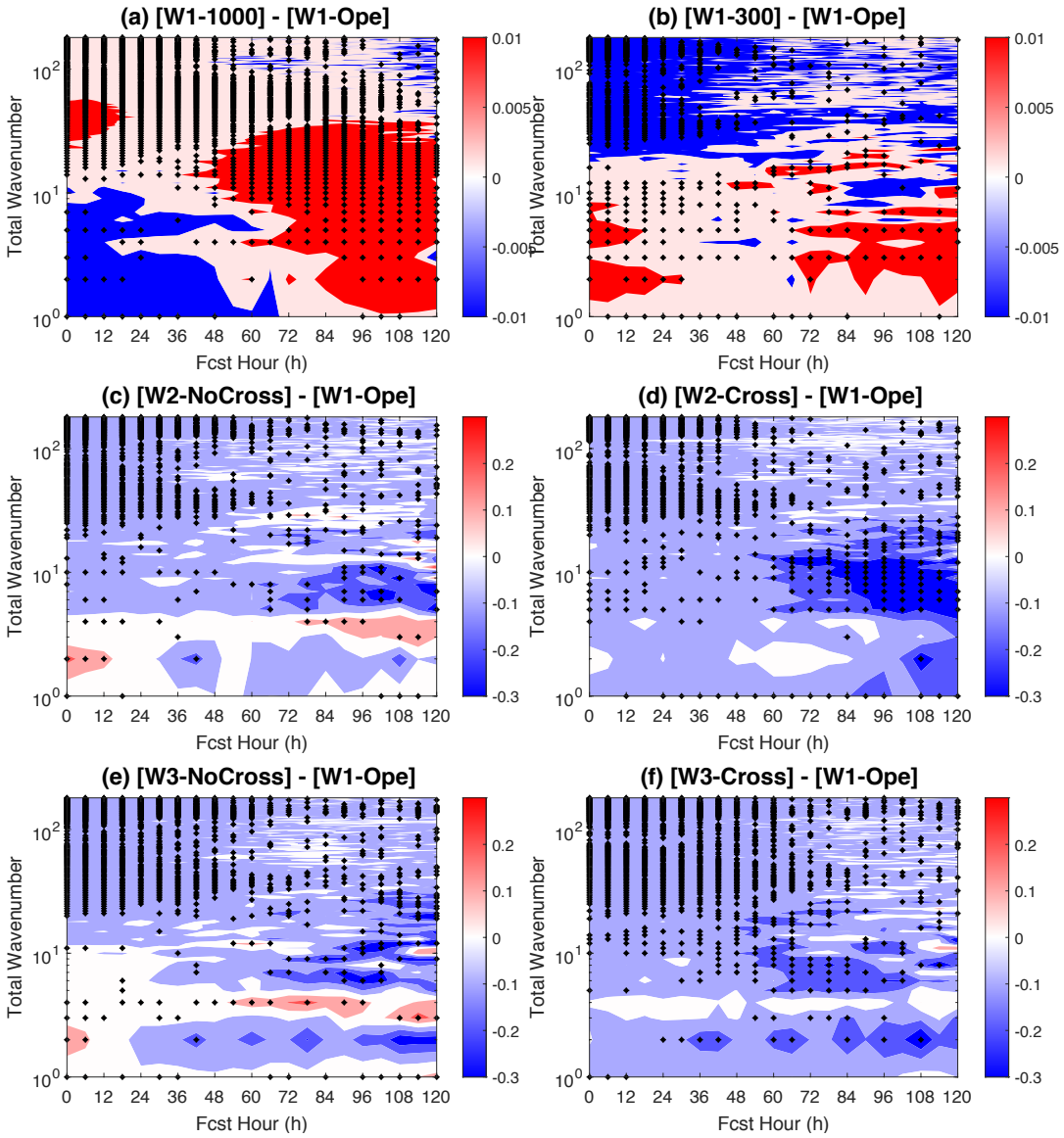

Figure 5.17 Power spectra of error total energy difference relative to W1-Ope in (a) W1-1000, (b) W1-300, (c) W2-NoCross, (d) W2-Cross, (e) W3-NoCross and (f) W3-Cross, as a function of forecast time to five days on the horizontal axis and total wavenumber on the vertical axis. See the texts for details of error total energy calculation. Blue (red) color indicates smaller (larger) error total energy relative to W1-Ope. The asterisk signs at the corresponding forecast times and total wavenumbers indicate that the difference from W1-Ope is statistically significant at or above 95% confidence level by applying the

paired permutation test with 1000 replicates combined with the FDR method at 95% confidence level.

5.7 Computational cost comparison

Table 5.2 summarizes the computational cost in wall clock time in each of the four components in a single 4DEnVar DA cycle. Further extending control variables in SDL increases the amount of computation in the 4DEnVar update and total computational cost, especially when a larger number of wavebands are adopted in SDL. In general, the ensemble forecasts are the most expensive component in an EnVar DA cycle. Therefore, the two-waveband and three-waveband SDL experiments only increase the total computational cost by 14% and 28%, respectively. Given that the two- and three-waveband SDL experiments significantly improve the global forecasts almost to five days over the strict reference W1-Ope and that W3-Cross even shows more accurate TC track forecasts than W1-Ope at shorter forecast lead times, SDL shows promises to be implemented operationally.

Table 5.2 Wall clock time in minutes for each of the four components in a single 4DEnVar DA cycle. The same number of 1260 cores were run in each component for different experiments

Expts	**Wall clock time in minutes in each of the four components in a single 4DEnvar DA cycle**				**Total wall clock time in minutes**	**Total cost ratio relative to W1**
	EnVar update	EnKF update	Control background Forecasts	Ensemble background forecasts		
W1	15	7	3	45	70	1.0
W2	25				80	1.14
W3	35				90	1.28

5.8 Conclusion and discussion

Two SDL variants, with (SDL-Cross) and without (SDL-NoCross) considering cross-waveband covariances were formulated based on the full B-preconditioned EnVar, and implemented in the GSI-based 4DEnVar system by further extending control variables. SDL performs a single-step simultaneous assimilation of all available observations, while applying different amount of localization to different scales of ensemble covariances. Complete removal of cross-waveband covariances in SDL-NoCross results in a local spatial averaging of ensemble covariances (Buehner and Charron 2007) and retains less heterogeneity of ensemble covariances (Caron et al. 2019). SDL-Cross that partially includes cross-waveband covariances retains more heterogeneity of ensemble covariances than SDL-NoCross. Performances of SDL-NoCross and SDL-Cross were evaluated for general global forecasts and TC track forecasts in the FV3-based GFS though one-month cycled DA experiments.

The two-waveband SDL experiments improve the global forecasts almost to five days over W1-1000 and W1-300 applying scale-invariant, level-invariant localization, and even over W1-Ope applying operationally-tuned, scale-invariant, level-dependent localization. By applying much wider horizontal localization length scale, W1-1000 generally degrades global forecasts below 50 hPa relative to W1-Ope. W1-300 applying tighter horizontal localization length scale shows worse global forecasts than W1-Ope at upper model levels especially for temperature forecasts. As for the TC track forecasts, the two-waveband SDL experiments outperform W1-1000 out to four days, and over W1-300 beyond three days. W1-Ope in general produces more accurate TC track forecasts than W1-1000 and W1-300.

To examine how SDL performs with respect to the number of decomposed wavebands, the three-waveband SDL experiments were compared with the two-waveband SDL experiments. In our design, the three-waveband SDL experiments apply tighter horizontal localization at medium-scale waveband than the two-waveband SDL experiments. This would lead to overall tighter effective horizontal localization compared to the two-waveband SDL experiments. As suggested in our cycled experiment results and the operationally-tuned, level-dependent, scale-invariant horizontal localization settings, wider (tighter) horizontal localization length scale is beneficial at upper (lower) model levels. Therefore, reduced overall effective horizontal localization length scale in the three-waveband SDL experiments may explain their general degraded (better) global forecasts above 50 hPa (below 100 hPa) than their two-waveband SDL counterparts. On the other hand, the degraded global forecasts of W3-Cross versus W2-Cross below 50 hPa at longer forecast lead times may be because the advantage of reduced overall effective horizontal localization length scale is overwhelmed by its more imbalance in the analysis. Compared to statistically indistinguishable TC track forecasts between W2-Cross and W1-Ope, W3-Cross shows significantly improved TC track forecasts than W1-Ope and W2-Cross within two days. This suggests the important role of scale separation in the SDL implementation especially for the TC track forecasts.

Due to the local spatial averaging of ensemble covariances, SDL-NoCross tends to show slightly better global forecasts than SDL-Cross at shorter forecast lead times. At longer forecast lead times, SDL-Cross outperforms SDL-NoCross for the global forecasts, especially in the two-waveband SDL experiments. However, comparable performances between SDL-NoCross and SDL-Cross were found in Caron et al. (2019) that run a 75-member ensemble populated by time-

lagged method (Van den Dool and Rukhovets 1994) in a regional 3DEnVar system. The better forecast performance of SDL-Cross than SDL-NoCross at longer forecast lead times in this study may benefit from more accurate estimate of cross-waveband covariances by directly running and updating an 80-member ensemble, retained more heterogeneity of ensemble covariances and more balanced analysis in SDL-Cross. The TC track forecasts between SDL-Cross and SDL-NoCross are generally consistent with their global wind forecasts at lower model levels. For example, W2-Cross shows significantly better TC track forecasts than W2-NoCross at longer forecast lead times.

Due to further extending control variables, the two- and three-waveband SDL experiments increase total computational cost by 14% and 28%, respectively, compared to the scale-invariant localization experiments. But SDL shows statistically significantly improved global forecasts and the potential of improving TC track forecasts over scale-invariant localization. Moreover, SDL without requiring additional ensemble forecasts is computationally much cheaper than directly increasing ensemble size. So SDL shows the potential to be implemented operationally.

In the current SDL experiments, the localization length scale is level-invariant. But the two-waveband SDL experiments showed increasing effective ensemble correlation length scales at large- and small-scale wavebands as model level increases (not shown here). Two-waveband level-dependent SDL experiments were motivated by increasing the horizontal localization length scales beyond 1000 km and 300 km *e*-folding distances for the large- and small-scale waveband, respectively, above 0.3 sigma model level following the increasing trend in W1-Ope (not shown here). However, these two-waveband level-dependent SDL experiments showed

degraded (comparable) global forecasts above (below) 200 hPa than the two-waveband level-invariant SDL experiments. The horizontal localization length scales in the two-waveband level-dependent SDL experiments need more tuning. Following Caron and Buehner (2018), the objective method proposed by Ménétrier et al. (2015) was attempted to determine optimal horizontal localization length scale at each waveband in our SDL experiments. But it showed worse global forecasts than our current SDL experiments. This may be related to the assumption of independent members in this objective method which may be not true in the operational NWP applications (Caron and Buehner 2018). Research on seeking the optimal effective localization length scale in SDL would be further explored in the future. In addition to applying SDL on the horizontal direction, further development of SDL on the vertical direction is ongoing in the GSI-based global 4DEnVar system.

Chapter 6: Summary and discussion

6.1 Summary

The ensemble-based DA, such as the EnKF and EnVar, has become popular in many operational NWP centers. Compared to the traditional variational DA that employs static background error covariances, the ensemble-based DA is able to provide flow-dependent background error information from running an ensemble of short-range forecasts. The ensemble-based DA is thus able to more realistically estimate the spatial, temporal and multivariate covariances, and further positively contribute to the analyses and subsequent forsecasts (Wang et al. 2007a, 2008a,b, 2009, 2013; Wang 2011; Lorenc et al. 2015; Kleist and Ide 2015a; Buehner et al. 2015). However, the computational cost in the ensemble-based DA is very expensive due to running an ensemble of forecasts, especially for the operational NWP applications. Constrained by the computational resources, the affordable ensemble size in the operational ensemble-based DA systems is much smaller than the degrees of freedom of the NWP models (Houtekamer and Zhang 2016). As a result, it will cause sampling error and rank-deficiency issue in the ensemble background covariances (Hamill 2006). This will further compromise the analysis accuracy. Direct increase of ensemble size is an ideal means to reduce sampling error, but the increased computational cost can be prohibitive. On the other hand, the covariance localization is a common strategy to ameliorate sampling error by reducing or removing the correlations with distant observations. Several popular localization methods are briefly reviewed in Chapter 2.

As sampling error will remain a major source of error in the ensemble-based DA, the dissertation covers three topics of efficient means to reduce sampling error and improve the ensemble background covariance estimate in the ensemble-based DA, by seeking cost-efficient

means to increase ensemble size as in Chapter 3 and improving the localization methods as in Chapters 4 and 5. Key results are summarized as follows.

Firstly, a valid-time-shifting (VTS) ensemble method is developed in Chapter 3 to inexpensively enlarge the background ensemble size in the GSI-based global hybrid 4DEnVar system operational at the NCEP. In VTS, the background ensembles initialized from the same previous cycle but valid at different lead times are used to populate the background ensemble at the analysis time. By utilizing the available ensemble forecasts, VTS is computationally much cheaper compared to directly increasing ensemble size. Two variants of VTS are further designed by directly shifting ensemble members (VTSM), or shifting ensemble perturbations (VTSP). Mathematically, VTSP is equivalent to applying a temporal smoothing of the ensemble background covariances, thus likely contributing to reducing sampling error. On the other hand, compared to the original background ensemble, VTSM increases total variances contributed by the original ensemble mean differences. In our cycled DA experiments using the GFS model at the NCEP, the background ensemble size is increased from 80 to 240 by applying VTSM or VTSP using different shifting time intervals (e.g., one, two and three hours). Overall, the cost in the VTSM and VTSP experiments only increases by 23%-27%. It is much cheaper compared to directly increasing ensemble size to 240 that almost triples the cost. Cycled DA experiment results show that VTSP is more efficient to improve global temperature and wind forecasts, especially in the two-hour VTSP experiment. Further diagnostics suggest that the improved global forecasts in VTSP could be associated with improved Gaussian distribution and correlation accuracy, and increased effective rank in the VTSP-populated background ensemble. Compared to using the original 80-member ensemble, VTSP and VTSM improve the tropical

storm track forecasts. Especially, the three-hour VTSM experiment even shows more accurate tropical storm track forecasts than directly increasing ensemble size to 240, likely owing to its better sampling of timing or phase errors in VTSM. These encouraging results further warrant potential application of VTSM and VTSP in the operational GSI-based global hybrid 4DEnVar system at the NCEP. In this Chapter, the significance test results in Figs. 3.4, 3.5, 3.6 and 3.13 adopting the paired *t*-test (which account for 35% of significance tests in this dissertation) are suboptimal due to the violation of distributional and independency assumptions that are required in the paired *t*-test. A more rigorous approach, such as the paired permutation test (Manly 2007) combined with the false discovery rate method (Wilks 2006) as in Chapter 5, is planned to investigate the significance of the results in Chapter 3 in the future.

Secondly, two distance-dependent localization methods, that is, the B-localization and R-localization methods, are mathematically compared in the generic EnKF context in Chapter 4. This contributes to understanding the fundamental difference of both localization methods. First, mathematical demonstration suggests that the B-localized background error covariances show higher rank than that applying the R-localization method. In addition, the mathematical demonstration also suggests a means of realizing the B-localization method through modulating and extending raw ensemble perturbations. This further motivates to implement the B-localization method in the ensemble transform Kalman filter (ETKF) that generally applies the R-localization method. The B-localized ETKF is referred to as the high-rank ETKF (HETKF) to distinguish the classic R-localized ETKF. Cycled DA experiments using the Lorenz Model II are further performed to evaluate and compare the performance of the HETKF and R-localized ETKF. The HETKF significantly and consistently improves the analysis accuracy compared to

the R-localized ETKF, especially for a small ensemble. Since the higher rank in the HETKF is expected to contribute more for a small ensemble, the improved analysis in the HETKF is likely associated with the high-rank in the B-localization method. In addition, the HETKF is less sensitive to the localization and inflation factors in contrast to the R-localized ETKF. This feature is especially attractive for the operational NWP applications.

Lastly, the simultaneous multi-scale DA capability is developed in the GSI-based global hybrid 4DEnVar system operational at the NCEP, by introducing scale-dependent-localization (SDL) with and without considering cross-waveband covariances (e.g., SDL-Cross and SDL-NoCross, respectively). SDL applies different amount of localization to different ranges of scales of background error covariances, while performing the single-step simultaneous assimilation of all the available observations. For example, stronger localization is generally applied to the small-scale background covariances. In this research, it first formulates SDL within the full-B preconditioned EnVar approach. SDL is then implemented in the GSI-based global hybrid 4DEnVar DA system, and evaluated for the general global forecasts and tropical storm track forecasts in the FV3-based GFS operational at the NCEP. Cycled DA experiments show that both SDL-NoCross and SDL-Cross improve the general global forecasts and tropical storm track forecasts compared to applying fixed uniform localization once at all scales and all model levels. An inter-comparison is further performed to evaluate the performance of SDL-NoCross and SDL-Cross that apply two and three wavebands. SDL-NoCross tends to outperform SDL-Cross at shorter forecast lead times. This may be owing to its better treatment of sampling errors that may result from the local spatial averaging of ensemble covariances by removing the cross-waveband covaraiances in SDL-NoCross. However, SDL-Cross shows more accurate general

global forecasts and tropical storm track forecasts at longer forecast lead times. This is likely associated with its retained higher degrees of heterogeneity of the ensemble covariances and more balanced analysis by including the cross-waveband covariances in SDL-Cross. In addition, the three-waveband SDL experiments that apply tighter horizontal localization at medium-scale waveband shows improved (degraded) global forecasts below (above) 50 hPa than their two-waveband SDL experiment counterparts, except that the better performance of the three-waveband SDL-Cross experiment over the two-waveband SDL-Cross experiment only lasts for three days. Finally, compared to the operational level-dependent fixed uniform localization at all scales, the two- and three-waveband SDL experiments improve the global forecasts to five days, and the three-waveband SDL-Cross experiment even shows more tropical storm track forecasts at shorter forecast lead times. These results show promises to implement SDL in the operational GSI-based global hybrid 4DEnVar system for the FV3-based GFS at the NCEP.

6.2 Discussion

It will remain difficult and challenging to run a very large ensemble especially for large-dimension operational NWP models in the near future. VTS with minimum cost increase could serve as an efficient alternative to reduce sampling error and improve background error covariance estimate in the ensemble-based DA. In contrast to utilizing the ensemble forecasts with the same initialization time but different valid times in VTS, the time-lagged ensemble method (Hoffman and Kalnay 1983) was also explored to populate the background ensemble by using the ensemble forecasts that are initialized from different previous analysis times but valid at the same analysis time. However, this time-lagged ensemble method only shows minimal or even negative impacts on the general global forecasts and tropical storm track forecasts in our

cycled DA experiments using the GSI-based global hybrid 4DEnVar system and the GFS model. This may be caused by less accurate representativeness of the true error in the ensemble forecasts at longer forecast times used in the time-lagged ensemble method. In addition, Lorenc (2017) further combined the VTS method and time-lagged ensemble method to populate the background ensemble, and found improved global forecasts over the stand-alone VTS or time-lagged ensemble method in a global hybrid 4DEnVar system. Given different performance of the individual VTS method and time-lagged ensemble method in our current study, it would be interesting to investigate the impacts of combining both methods in the GSI-based global hybrid 4DEnVar system. Furthermore, different performance of VTSM and VTSP for the global forecasts and tropical storm track forecasts in a global modeling system in our current study inspires to evaluate their impacts in a regional model system in the future.

The current computational constraint makes covariance localization required for successful application of the ensemble-based DA especially for operational NWP applications. The current implementation of the HETKF, as discussed in Chapter 4, has the issue with regard to more robustly sub-selecting analysis perturbations for next cycle. This issue is resolved in the gain form of the ETKF (GETKF) in Bishop et al. (2017). In the GETKF, the analysis perturbations are selected to be mathematically more consistent with the modulated background ensembles. Our further examination shows that the HETKF with the deterministic perturbation sub-selection approach performs comparably as the GETKF when the full model-space localization matrix is isotropic. For a non-isotropic model-space localization matrix, however, the GETKF produces more accurate analysis (Bishop et al. 2017). In addition, the HETKF and GETKF apply the model-space localization through modulating raw ensemble perturbations, in

contrast to the popular operational EnSRF and LETKF that perform the localization on the observation space. For the assimilation of integral-type observations (e.g., satellite radiances and radar reflectivity) whose physical location is difficult to be explicitly defined, it is expected that the HETKF and GETKF with the model-space localization would be more advantageous compared to the EnSRF and LETKF. Therefore, future implementation of the HETKF or GETKF is very appealing for operational NWP applications.

The future global model will be able to resolve wider ranges of scales due to increasing computing power. The multi-scale DA capability will become necessary and play an essential role. In our current study, SDL is applied on the horizontal localization. In the future, it will be interesting to examine the impacts of further extending SDL on the vertical localization. In addition, a multi-resolution ensemble capability was developed in the most recent in the GSI-based global hybrid 4DEnVar system (Kay and Wang 2020) by mixing the background ensembles of different resolutions. Impacts of further combining the SDL capability and the multi-resolution ensemble capability will be investigated in the GSI-based global hybrid 4DEnVar system. Furthermore, the implementation of SDL by further extending the control variables inspires to develop and implement the localization method that varies with the model variable, time and location.

Finally, the above three research topics could be further extended to other applications and combined to provide complementary benefits of reducing sampling error and improving background covariance estimate in the ensemble-based DA. With minimum modifications, for example, VTS can be easily implemented in the popular operational EnKF systems that adopt the EnSRF and LETKF algorithm. In addition, the VTS and time-lagged ensemble methods take

advantage of ensemble forecasts of different forecast lengths to populate the background ensemble. These ensemble forecasts of different forecast lengths may contain richer information of forecast error at different scales. Further combining SDL with the VTS or time-lagged ensemble method shall be straightforward and expected to provide additional benefits. In the HETKF and GETKF, the modes-space B-localization is realized by modulating raw ensemble perturbations. This inspires to develop SDL in the EnKF approach by modulating the raw ensemble perturbations at different scales with different modulation functions that represent different amount of localization. This SDL capability in the EnKF approach could also be further combined with the VTS and time-lagged methods.

Bibliography

Anderson, J., and L. Lei, 2013: Empirical Localization of Observation Impact in Ensemble Kalman Filters. *Mon. Wea. Rev.*, **141**, 4140–4153, doi:10.1175/MWR-D-12-00330.1.

Anderson, J. L., 2001: An Ensemble Adjustment Kalman Filter for Data Assimilation. *Mon. Wea. Rev.*, **129**, 2884–2903, doi:10.1175/1520-0493(2001)129<2884:AEAKFF>2.0.CO;2.

——, 2007: Exploring the need for localization in ensemble data assimilation using a hierarchical ensemble filter. *Physica. D* , **230**, 99–111, doi:10.1016/J.PHYSD.2006.02.011.

Bishop, C., and D. Hodyss, 2009a: Ensemble covariances adaptively localized with ECO-RAP. Part 1: Tests on simple error models. *Tellus A Dyn. Meteorol. Oceanogr.*, **61**, 84–96.

——, and ——, 2009b: Ensemble covariances adaptively localized with ECO-RAP. Part 2: A strategy for the atmosphere. *Tellus A Dyn. Meteorol. Oceanogr.*, **61**, 97–111.

Bishop, C. H., and D. Hodyss, 2007: Flow-adaptive moderation of spurious ensemble correlations and its use in ensemble-based data assimilation. *Quart. J. Roy. Meteor. Soc.*, **133**, 2029–2044, doi:10.1002/qj.169.

——, and ——, 2009c: Ensemble covariances adaptively localized with ECO-RAP. Part 2: A strategy for the atmosphere. *Tellus*, **61 A**, 97–111, doi:10.1111/j.1600-0870.2008.00372.x.

——, B. J. Etherton, S. J. Majumdar, C. H. Bishop, B. J. Etherton, and S. J. Majumdar, 2001: Adaptive Sampling with the Ensemble Transform Kalman Filter. Part I: Theoretical Aspects. *Mon. Wea. Rev.*, **129**, 420–436, doi:10.1175/1520-0493(2001)129<0420:ASWTET>2.0.CO;2.

——, B. Huang, and X. Wang, 2015: A Non-Variational Consistent Hybrid Ensemble Filter. *Mon. Wea. Rev.*, 150723141240007, doi:10.1175/MWR-D-14-00391.1.

——, J. S. Whitaker, and L. Lei, 2017: Gain Form of the Ensemble Transform Kalman Filter and Its Relevance to Satellite Data Assimilation with Model Space Ensemble Covariance Localization. *Mon. Wea. Rev.*, **145**, 4575–4592, doi:10.1175/MWR-D-17-0102.1.

Bonavita, M., L. Isaksen, and E. Hólm, 2012: On the use of EDA background error variances in the ECMWF 4D-Var. *Quart. J. Roy. Meteor. Soc.*, **138**, 1540–1559, doi:10.1002/qj.1899.

——, E. Hólm, L. Isaksen, and M. Fisher, 2016: The evolution of the ECMWF hybrid data assimilation system. *Quart. J. Roy. Meteor. Soc.*, **142**, 287–303, doi:10.1002/qj.2652.

Bowler, N. E., and Coauthors, 2017: Inflation and localization tests in the development of an ensemble of 4D-ensemble variational assimilations. *Quart. J. Roy. Meteor. Soc.*, **143**, 1280–1302, doi:10.1002/qj.3004.

Buehner, M., 2005: Ensemble-derived stationary and flow-dependent background-error covariances: Evaluation in a quasi-operational NWP setting. *Quart. J. Roy. Meteor. Soc.*, **131**, 1013–1043, doi:10.1256/qj.04.15.

——, 2012a: Evaluation of a spatial/spectral covariance localization approach for atmospheric

data assimilation. *Mon. Wea. Rev.*, **140**, 617–636, doi:10.1175/MWR-D-10-05052.1.

——, 2012b: Evaluation of a Spatial/Spectral Covariance Localization Approach for Atmospheric Data Assimilation. *Mon. Wea. Rev.*, **140**, 617–636, doi:10.1175/MWR-D-10-05052.1.

——, and M. Charron, 2007: Spectral and spatial localization of background-error correlations for data assimilation. *Quart. J. Roy. Meteor. Soc.*, **133**, 615–630, doi:10.1002/qj.50.

——, and A. Shlyaeva, 2015: Scale-dependent background-error covariance localisation. *Tellus, Ser. A Dyn. Meteorol. Oceanogr.*, **6**, doi:10.3402/tellusa.v67.28027.

——, and Coauthors, 2010: Intercomparison of Variational Data Assimilation and the Ensemble Kalman Filter for Global Deterministic NWP. Part I: Description and Single-Observation Experiments. *Mon. Wea. Rev.*, **138**, 1550–1566, doi:10.1175/2009MWR3157.1.

Buehner, M., J. Morneau, and C. Charette, 2013: Four-dimensional ensemble-variational data assimilation for global deterministic weather prediction. *Nonlinear Processes Geophys.*, **20**, 669–682, doi:10.5194/npg-20-669-2013.

Buehner, M., and Coauthors, 2015: Implementation of Deterministic Weather Forecasting Systems Based on Ensemble–Variational Data Assimilation at Environment Canada. Part I: The Global System. *Mon. Wea. Rev.*, **143**, 2532–2559, doi:10.1175/MWR-D-14-00354.1.

Campbell, W. F., C. H. Bishop, and D. Hodyss, 2010: Vertical Covariance Localization for Satellite Radiances in Ensemble Kalman Filters. *Mon. Wea. Rev.*, **138**, 282–290, doi:10.1175/2009MWR3017.1.

Caron, J., É. Arbogast, and M. Buehner, Scale-dependent localization in ensemble- variational data assimilation : Application in global and convective-scale systems.

Caron, J. F., and M. Buehner, 2018: Scale-dependent background error covariance localization: Evaluation in a global deterministic weather forecasting system. *Mon. Wea. Rev.*, **146**, 1367–1381, doi:10.1175/MWR-D-17-0369.1.

——, Y. Michel, T. Montmerle, and É. Arbogast, 2019: Improving background error covariances in a 3D ensemble-variational data assimilation system for regional NWP. *Mon. Wea. Rev.*, **147**, 135–151, doi:10.1175/MWR-D-18-0248.1.

Chen, J. H., and Coauthors, 2019: Advancements in Hurricane Prediction With NOAA's Next-Generation Forecast System. *Geophys. Res. Lett.*, **46**, 4495–4501, doi:10.1029/2019GL082410.

Clayton, A. M., A. C. Lorenc, and D. M. Barker, 2013: Operational implementation of a hybrid ensemble/4D-Var global data assimilation system at the Met Office. *Quart. J. Roy. Meteor. Soc.*, **139**, 1445–1461, doi:10.1002/qj.2054.

Dee, D. P., and Coauthors, 2011: The ERA-Interim reanalysis: Configuration and performance of the data assimilation system. *Quart. J. Roy. Meteor. Soc.*, **137**, 553–597.

Van den Dool, H. M., and L. Rukhovets, 1994: On the Weights for an Ensemble-Averaged 6–10-

Day Forecast. *Wea. Forecasting*, **9**, 457–465, doi:10.1175/1520-0434(1994)009<0457:OTWFAE>2.0.CO;2.

Environmental Modeling Center, 2020: The Global Forecast System (GFS) - Global Spectral Model (GSM). Available at https://www.emc.ncep.noaa.gov/emc/pages/numerical_forecast_systems/gfs/documentation.php.

European Centre for Medium-Range Weather Forecasts, 2020: ERA iNTERIM, Daily. Available at https://apps.ecmwf.int/datasets/data/interim-full-daily/levtype=sfc/.

Evensen, G., 1994: Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics. *J. Geophys. Res.*, **99**, 10143, doi:10.1029/94JC00572.

Fairbairn, D., S. R. Pring, A. C. Lorenc, and I. Roulstone, 2014: A comparison of 4DVar with ensemble data assimilation methods. *Quart. J. Roy. Meteor. Soc.*, **140**, 281–294, doi:10.1002/qj.2135.

Fisher, M., 2003: Background error covariance modelling. *Seminar on Recent Development in Data Assimilation for Atmosphere and Ocean*, 45–63.

Gaspari, G., and S. E. Cohn, 1999: Construction of correlation functions in two and three dimensions. *Quart. J. Roy. Meteor. Soc.*, **125**, 723–757, doi:10.1002/qj.49712555417.

Gasperoni, N. A., and X. Wang, 2015: Adaptive localization for the ensemble-based observation impact estimate using regression confidence factors. *Mon. Wea. Rev.*, **143**, 1981–2000, doi:10.1175/MWR-D-14-00272.1.

Ghil, M., and P. Malanotte-Rizzoli, 1991: Data assimilation in meteorology and oceanography. *Advances in geophysics*, Vol. 33 of, Elsevier, 141–266.

Greybush, S. J., E. Kalnay, T. Miyoshi, K. Ide, and B. R. Hunt, 2011: Balance and Ensemble Kalman Filter Localization Techniques. *Mon. Wea. Rev.*, **139**, 511–522, doi:10.1175/2010MWR3328.1.

Gustafsson, N., J. Bojarova, and O. Vignes, 2014: A hybrid variational ensemble data assimilation for the HIgh Resolution Limited Area Model (HIRLAM). *Nonlinear Processes Geophys.*, **21**, 303–323, doi:10.5194/npg-21-303-2014.

Hamill, T. M., 2001: Interpretation of Rank Histograms for Verifying Ensemble Forecasts. *Mon. Wea. Rev.*, **129**, 550–560, doi:10.1175/1520-0493(2001)129<0550:IORHFV>2.0.CO;2.

——, 2006: Ensemble-based atmospheric data assimilation. *Predictability of Weather and Climate*, T. Palmer and R. Hagedorn, Eds., Cambridge University Press, Cambridge, 124–156.

——, and C. Snyder, 2000: A Hybrid Ensemble Kalman Filter–3D Variational Analysis Scheme. *Mon. Wea. Rev.*, **128**, 2905–2919, doi:10.1175/1520-0493(2000)128<2905:AHEKFV>2.0.CO;2.

Harris, L. M., and S.-J. Lin, 2013: A two-way nested global-regional dynamical core on the

cubed-sphere grid. *Mon. Wea. Rev.*, **141**, 283–306.
Hoffman, R. N., and E. Kalnay, 1983: Lagged average forecasting, an alternative to Monte Carlo forecasting. *Tellus A*, **35 A**, 100–118, doi:10.1111/j.1600-0870.1983.tb00189.x.
Holland, B., and X. Wang, 2013: Effects of sequential or simultaneous assimilation of observations and localization methods on the performance of the ensemble Kalman filter. *Quart. J. Roy. Meteor. Soc.*, **139**, 758–770, doi:10.1002/qj.2006.
Hollingsworth, A., and P. Lönnberg, 1986: The statistical structure of short-range forecast errors as determined from radiosonde data. Part I: The wind field. *Tellus A*, **38**, 111–136.
Houtekamer, P. L., and H. L. Mitchell, 1998: Data Assimilation Using an Ensemble Kalman Filter Technique. *Mon. Wea. Rev.*, **126**, 796–811, doi:10.1175/1520-0493(1998)126<0796:DAUAEK>2.0.CO;2.
Houtekamer, P. L., and H. L. Mitchell, 2001: A Sequential Ensemble Kalman Filter for Atmospheric Data Assimilation. *Mon. Wea. Rev.*, **129**, 123–137, doi:10.1175/1520-0493(2001)129<0123:ASEKFF>2.0.CO;2.
——, and ——, 2005: Ensemble Kalman filtering. *Quart. J. Roy. Meteor. Soc.*, **131**, 3269–3289, doi:10.1256/qj.05.135.
Houtekamer, P. L., and F. Zhang, 2016: Review of the Ensemble Kalman Filter for Atmospheric Data Assimilation. *Mon. Wea. Rev.*, **144**, 4489–4532, doi:10.1175/MWR-D-15-0440.1.
——, X. Deng, H. L. Mitchell, S.-J. Baek, and N. Gagnon, 2014: Higher Resolution in an Operational Ensemble Kalman Filter. *Mon. Wea. Rev.*, **142**, 1143–1162, doi:10.1175/MWR-D-13-00138.1.
Huang, B., and X. Wang, 2018a: On the Use of Cost-Effective Valid-Time-Shifting (VTS) Method to Increase Ensemble Size in the GFS Hybrid 4DEnVar System. *Mon. Wea. Rev.*, **146**, 2973–2998, doi:10.1175/MWR-D-18-0009.1.
——, and ——, 2018b: On the Use of Cost-Effective Valid-Time-Shifting (VTS) Method to Increase Ensemble Size in the GFS Hybrid 4DEnVar System. *Mon. Wea. Rev.*, **146**, 2973–2998, doi:10.1175/MWR-D-18-0009.1.
——, ——, and C. H. Bishop, 2019: The High-Rank Ensemble Transform Kalman Filter. *Mon. Wea. Rev.*, **147**, 3025–3043, doi:10.1175/MWR-D-18-0210.1.
Hunt, B. R., E. J. Kostelich, and I. Szunyogh, 2007: Efficient data assimilation for spatiotemporal chaos: A local ensemble transform Kalman filter. *Physica. D* , **230**, 112–126, doi:10.1016/j.physd.2006.11.008.
Ide, K., P. Courtier, M. Ghil, and A. C. Lorenc, 1997: Uniied Notation for Data Assimilation: Operational, Sequential and Variational. *J. Met. Soc. Japan*, **75**, 181–189.
Janjić, T., L. Nerger, A. Albertella, J. Schröter, and S. Skachko, 2011: On Domain Localization in Ensemble-Based Kalman Filter Algorithms. *Mon. Wea. Rev.*, **139**, 2046–2060, doi:10.1175/2011MWR3552.1.

Julier, S. J., and J. K. Uhlmann, 1997: New extension of the Kalman filter to nonlinear systems. *Signal processing, sensor fusion, and target recognition VI*, Vol. 3068 of, International Society for Optics and Photonics, 182–193.

Kalman, R. E., and R. S. Bucy, 1961: New Results in Linear Filtering and Prediction Theory. *J. Basic Eng.*, **83**, 95, doi:10.1115/1.3658902.

Kalnay, E., 2003: *Atmospheric modeling, data assimilation and predictability*. Cambridge university press,.

Kang, J., E. Kalnay, J. Liu, I. Fung, T. Miyoshi, and K. Ide, 2011: "Variable localization" in an ensemble Kalman filter: Application to the carbon cycle data assimilation. *J. Geophys. Res. Atmos.*, **116**.

Kepert, J. D., 2009: Covariance localisation and balance in an ensemble Kalman filter. *Quart. J. Roy. Meteor. Soc.*, **135**, 1157–1176.

Kleist, D. T., and K. Ide, 2015a: An OSSE-Based Evaluation of Hybrid Variational–Ensemble Data Assimilation for the NCEP GFS. Part II: 4DEnVar and Hybrid Variants. *Mon. Wea. Rev.*, **143**, 452–470, doi:10.1175/MWR-D-13-00350.1.

——, and ——, 2015b: 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.

——, D. F. Parrish, J. C. Derber, R. Treadon, W.-S. Wu, and S. Lord, 2009: Introduction of the GSI into the NCEP Global Data Assimilation System. *Wea. Forecasting*, **24**, 1691–1705, doi:10.1175/2009WAF2222201.1.

Kretschmer, M., B. R. Hunt, and E. Ott, 2015: Data assimilation using a climatologically augmented local ensemble transform Kalman filter. *Tellus, Ser. A Dyn. Meteorol. Oceanogr.*, **67**, 1–9, doi:10.3402/tellusa.v67.26617.

Kuhl, D., and Coauthors, 2007: Assessing Predictability with a Local Ensemble Kalman Filter. *J. Atmos. Sci.*, **64**, 1116–1140, doi:10.1175/JAS3885.1.

Kuhl, D. D., T. E. Rosmond, C. H. Bishop, J. McLay, and N. L. Baker, 2013: Comparison of Hybrid Ensemble/4DVar and 4DVar within the NAVDAS-AR Data Assimilation Framework. *Mon. Wea. Rev.*, **141**, 2740–2758, doi:Doi 10.1175/Mwr-D-12-00182.1.

Kutty, G., and X. Wang, 2015: A Comparison of the Impacts of Radiosonde and AMSU Radiance Observations in GSI Based 3DEnsVar and 3DVar Data Assimilation Systems for NCEP GFS. *Adv. Meteor.*, **2015**, 1–17, doi:10.1155/2015/280546.

Laloyaux, P., M. Balmaseda, D. Dee, K. Mogensen, and P. Janssen, 2016: A coupled data assimilation system for climate reanalysis. *Quart. J. Roy. Meteor. Soc.*, **142**, 65–78.

Lei, L., and J. S. Whitaker, 2016: A Four-Dimensional Incremental Analysis Update for the Ensemble Kalman Filter. *Mon. Wea. Rev.*, **144**, 2605–2621, doi:10.1175/MWR-D-15-0246.1.

——, and ——, 2017: Evaluating the trade-offs between ensemble size and ensemble resolution in an ensemble-variational data assimilation system. *J. Adv. Model. Earth Syst.*, **9**, 781–789, doi:10.1002/2016MS000864.

Lewis, J. M., S. Lakshmivarahan, and S. Dhall, 2006: *Dynamic data assimilation: a least squares approach*. Cambridge University Press,.

Lin, S.-J., 2004: A "vertically Lagrangian" finite-volume dynamical core for global models. *Mon. Wea. Rev.*, **132**, 2293–2307.

Lorenc, A. C., 1986: Analysis methods for numerical weather prediction. *Quart. J. Roy. Meteor. Soc.*, **112**, 1177–1194.

——, 2003a: The potential of the ensemble Kalman filter for NWP—a comparison with 4D-Var. *Q. J. R. Meteorol. Soc. A J. Atmos. Sci. Appl. Meteorol. Phys. Oceanogr.*, **129**, 3183–3203.

Lorenc, A. C., 2003b: The potential of the ensemble Kalman filter for NWP—a comparison with 4D-Var. *Quart. J. Roy. Meteor. Soc.*, **129**, 3183–3203, doi:10.1256/qj.02.132.

——, 2017a: Improving ensemble covariances in hybrid variational data assimilation without increasing ensemble size. *Quart. J. Roy. Meteor. Soc.*, **143**, 1062–1072, doi:10.1002/qj.2990.

——, 2017b: Improving ensemble covariances in hybrid variational data assimilation without increasing ensemble size. *Quart. J. Roy. Meteor. Soc.*, **143**, 1062–1072, doi:10.1002/qj.2990.

Lorenc, A. C., and Coauthors, 2000: The Met. Office global three-dimensional variational data assimilation scheme. *Quart. J. Roy. Meteor. Soc.*, **126**, 2991–3012.

Lorenc, A. C., N. E. Bowler, A. M. Clayton, S. R. Pring, and D. Fairbairn, 2015: Comparison of Hybrid-4DEnVar and Hybrid-4DVar Data Assimilation Methods for Global NWP. *Mon. Wea. Rev.*, **143**, 212–229, doi:10.1175/MWR-D-14-00195.1.

Lorenz, E. N., 1996: Predictability: a problem partly solved. *Proc. Seminar on Predictability*, Shinfield Park, Reading, United Kingdom, European Centre for Medium-Range Weather Forecasts, 1–18.

——, 2005: Designing Chaotic Models. *J. Atmos. Sci.*, **62**, 1574–1587, doi:10.1175/JAS3430.1.

Lu, C., H. Yuan, B. E. Schwartz, and S. G. Benjamin, 2007: Short-Range Numerical Weather Prediction Using Time-Lagged Ensembles. *Wea. Forecasting*, **22**, 580–595, doi:10.1175/WAF999.1.

Lu, H., Q. Xu, M. Yao, and S. Gao, 2011: Time-expanded sampling for ensemble-based filters: Assimilation experiments with real radar observations. *Adv. Atmos. Sci.*, **28**, 743–757, doi:10.1007/s00376-010-0021-4.

Lynch, P., and X.-Y. Huang, 1992: Initialization of the HIRLAM Model Using a Digital Filter. *Mon. Wea. Rev.*, **120**, 1019–1034, doi:10.1175/1520-0493(1992)120<1019:IOTHMU>2.0.CO;2.

Majumdar, S. J., C. H. Bishop, B. J. Etherton, Z. Toth, S. J. Majumdar, C. H. Bishop, B. J. Etherton, and Z. Toth, 2002: Adaptive Sampling with the Ensemble Transform Kalman Filter. Part II: Field Program Implementation. *Mon. Wea. Rev.*, **130**, 1356–1369, doi:10.1175/1520-0493(2002)130<1356:ASWTET>2.0.CO;2.

Manly, B. F. J., 2006: *Randomization, bootstrap and Monte Carlo methods in biology*. CRC press,.

Ménétrier, B., T. Montmerle, Y. Michel, and L. Berre, 2015: Linear filtering of sample covariances for ensemble-based data assimilation. Part I: Optimality criteria and application to variance filtering and covariance localization. *Mon. Wea. Rev.*, **143**, 1622–1643, doi:10.1175/MWR-D-14-00157.1.

Miyoshi, T., and S. Yamane, 2007: Local Ensemble Transform Kalman Filtering with an AGCM at a T159/L48 Resolution. *Mon. Wea. Rev.*, **135**, 3841–3861, doi:10.1175/2007MWR1873.1.

——, and K. Kondo, 2013: A multi-scale localization approach to an ensemble Kalman filter. *Sci. Online Lett. Atmos.*, **9**, 170–173, doi:10.2151/sola.2013-038.

——, ——, and T. Imamura, 2014: The 10,240-member ensemble Kalman filtering with an intermediate AGCM. *Geophys. Res. Lett.*, **41**, 5264–5271, doi:10.1002/2014GL060863.

National Hurricane Center and Central pacific Hurricane Center, 2020: Tropical Cyclone Reports. Available at https://www.nhc.noaa.gov/data/#hurdat.

National Weather Service, 2020:Phase 2 : Continued Dynamical Core Computational Performance Evaluation. Available at https://www.weather.gov/sti/stimodeling_nggps_implementation_atmdynamics.

Naval Oceanography Portal, 2020: Best Track Archive. Available at https://www.metoc.navy.mil/jtwc/jtwc.html?best-tracks.

Nerger, L., T. Janjić, J. Schröter, and W. Hiller, 2012: A regulated localization scheme for ensemble-based Kalman filters. *Quart. J. Roy. Meteor. Soc.*, **138**, 802–812, doi:10.1002/qj.945.

Oczkowski, M., I. Szunyogh, and D. J. Patil, 2005: Mechanisms for the Development of Locally Low-Dimensional Atmospheric Dynamics. *J. Atmos. Sci.*, **62**, 1135–1156, doi:10.1175/JAS3403.1.

Palmer, T. N., R. Gelaro, J. Barkmeijer, R. Buizza, T. N. Palmer, R. Gelaro, J. Barkmeijer, and R. Buizza, 1998: Singular Vectors, Metrics, and Adaptive Observations. *J. Atmos. Sci.*, **55**, 633–653, doi:10.1175/1520-0469(1998)055<0633:SVMAAO>2.0.CO;2.

Palmer, T. N., R. Buizza, F. Doblas-Reyes, T. Jung, M. Leutbecher, G. J. Shutts, M. Steinheimer, and A. Weisheimer, 2009: Stochastic Parametrization and Model Uncertainty.

Pan, Y., K. Zhu, M. Xue, X. Wang, M. Hu, S. G. Benjamin, S. S. Weygandt, and J. S. Whitaker, 2014: A GSI-based coupled EnSRF–En3DVar hybrid data assimilation system for the operational rapid refresh model: Tests at a reduced resolution. *Mon. Wea. Rev.*, **142**, 3756–

3780.
Parrish, D. F., and J. C. Derber, 1992: The National Meteorological Center's spectral statistical-interpolation analysis system. *Mon. Wea. Rev.*, **120**, 1747–1763.
Patil, D. J., B. R. Hunt, E. Kalnay, J. A. Yorke, and E. Ott, 2001: Local low dimensionality of atmospheric dynamics. *Phys. Rev. Lett.*, **86**, 5878–5881, doi:10.1103/PhysRevLett.86.5878.
Rainwater, S., and B. Hunt, 2013: Mixed Resolution Ensemble Data Assimilation. *Mon. Wea. Rev.*, 130305130008003, doi:10.1175/MWR-D-12-00234.1.
Rawlins, F., S. P. Ballard, K. J. Bovis, A. M. Clayton, D. Li, G. W. Inverarity, A. C. Lorenc, and T. J. Payne, 2007: The Met Office global four-dimensional variational data assimilation scheme. *Q. J. R. Meteorol. Soc. A J. Atmos. Sci. Appl. Meteorol. Phys. Oceanogr.*, **133**, 347–362.
Raynaud, L., L. Berre, and G. Desroziers, 2008: Spatial averaging of ensemble-based background-error variances. *Quart. J. Roy. Meteor. Soc.*, **134**, 1003–1014, doi:10.1002/qj.245.
——, ——, and ——, 2009: Objective filtering of ensemble-based background-error variances. *Quart. J. Roy. Meteor. Soc.*, **135**, 1177–1199, doi:10.1002/qj.438.
Sakov, P., and L. Bertino, 2011: Relation between two common localisation methods for the EnKF. *Comput. Geosci.*, **15**, 225–237, doi:10.1007/s10596-010-9202-6.
Straus, D., and D. Paolino, 2008: Intermediate time error growth and predictability: tropics versus mid-latitudes. *Tellus*, **61A**, 579–586, doi:10.1111/j.1600-0870.2009.00411.x.
Tribbia, J. J., and D. P. Baumhefner, 2004: Scale interactions and atmospheric predictability: An updated perspective. *Mon. Wea. Rev.*, **132**, 703–713.
Wang, X., 2010a: Incorporating Ensemble Covariance in the Gridpoint Statistical Interpolation Variational Minimization: A Mathematical Framework. *Mon. Wea. Rev.*, **138**, 2990–2995, doi:10.1175/2010MWR3245.1.
——, 2010b: Incorporating ensemble covariance in the gridpoint statistical interpolation variational minimization: A mathematical framework. *Mon. Wea. Rev.*, **138**, 2990–2995, doi:10.1175/2010MWR3245.1.
——, 2011: Application of the WRF Hybrid ETKF–3DVAR Data Assimilation System for Hurricane Track Forecasts. *Wea. Forecasting*, **26**, 868–884, doi:10.1175/WAF-D-10-05058.1.
——, and C. H. Bishop, 2003: A Comparison of Breeding and Ensemble Transform Kalman Filter Ensemble Forecast Schemes. *J. Atmos. Sci.*, **60**, 1140–1158, doi:10.1175/1520-0469(2003)060<1140:ACOBAE>2.0.CO;2.
——, and C. H. Bishop, 2005: Improvement of ensemble reliability with a new dressing kernel. *Q. J. R. Meteorol. Soc. A J. Atmos. Sci. Appl. Meteorol. Phys. Oceanogr.*, **131**, 965–986.
——, and T. Lei, 2014: GSI-Based Four-Dimensional Ensemble–Variational (4DEnsVar) Data

Assimilation: Formulation and Single-Resolution Experiments with Real Data for NCEP Global Forecast System. *Mon. Wea. Rev.*, **142**, 3303–3325, doi:10.1175/MWR-D-13-00303.1.

——, C. H. Bishop, S. J. Julier, X. Wang, C. H. Bishop, and S. J. Julier, 2004: Which Is Better, an Ensemble of Positive–Negative Pairs or a Centered Spherical Simplex Ensemble? *Mon. Wea. Rev.*, **132**, 1590–1605, doi:10.1175/1520-0493(2004)132<1590:WIBAEO>2.0.CO;2.

——, T. M. Hamill, J. S. Whitaker, C. H. Bishop, X. Wang, T. M. Hamill, J. S. Whitaker, and C. H. Bishop, 2007a: A Comparison of Hybrid Ensemble Transform Kalman Filter–Optimum Interpolation and Ensemble Square Root Filter Analysis Schemes. *Mon. Wea. Rev.*, **135**, 1055–1076, doi:10.1175/MWR3307.1.

——, C. Snyder, and T. M. Hamill, 2007b: On the theoretical equivalence of differently proposed ensemble - 3DVAR hybrid analysis schemes. *Mon. Wea. Rev.*, **135**, 222–227, doi:10.1175/MWR3282.1.

——, ——, ——, X. Wang, C. Snyder, and T. M. Hamill, 2007c: On the Theoretical Equivalence of Differently Proposed Ensemble–3DVAR Hybrid Analysis Schemes. *Mon. Wea. Rev.*, **135**, 222–227, doi:10.1175/MWR3282.1.

——, D. M. Barker, C. Snyder, and T. M. Hamill, 2008a: A hybrid ETKF-3DVAR data assimilation scheme for the WRF model. Part I: Observation system simulation experiment. *Mon. Wea. Rev.*, **136**, 5116–5131, doi:10.1175/2008MWR2444.1.

——, ——, ——, ——, X. Wang, D. M. Barker, C. Snyder, and T. M. Hamill, 2008b: A Hybrid ETKF–3DVAR Data Assimilation Scheme for the WRF Model. Part I: Observing System Simulation Experiment. *Mon. Wea. Rev.*, **136**, 5116–5131, doi:10.1175/2008MWR2444.1.

——, ——, ——, ——, ——, ——, ——, and ——, 2008c: A Hybrid ETKF–3DVAR Data Assimilation Scheme for the WRF Model. Part II: Real Observation Experiments. *Mon. Wea. Rev.*, **136**, 5132–5147, doi:10.1175/2008MWR2445.1.

——, T. M. Hamill, J. S. Whitaker, C. H. Bishop, X. Wang, T. M. Hamill, J. S. Whitaker, and C. H. Bishop, 2009: A Comparison of the Hybrid and EnSRF Analysis Schemes in the Presence of Model Errors due to Unresolved Scales. *Mon. Wea. Rev.*, **137**, 3219–3232, doi:10.1175/2009MWR2923.1.

——, 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, doi:10.1175/MWR-D-12-00141.1.

Whitaker, J. S., and T. M. Hamill, 2002: Ensemble Data Assimilation without Perturbed Observations. *Mon. Wea. Rev.*, **130**, 1913–1924, doi:10.1175/1520-0493(2002)130<1913:EDAWPO>2.0.CO;2.

——, ——, X. Wei, Y. Song, and Z. Toth, 2008: Ensemble Data Assimilation with the NCEP Global Forecast System. *Mon. Wea. Rev.*, **136**, 463–482, doi:10.1175/2007MWR2018.1.

——, ——, J. S. Whitaker, and T. M. Hamill, 2012: Evaluating Methods to Account for System Errors in Ensemble Data Assimilation. *Mon. Wea. Rev.*, **140**, 3078–3089, doi:10.1175/MWR-D-11-00276.1.

Wilks, D. S., 2006: On "field significance" and the false discovery rate. *J. Appl. Meteorol. Climatol.*, **45**, 1181–1189.

Xu, Q., L. Wei, H. Lu, C. Qiu, and Q. Zhao, 2008: Time-expanded sampling for ensemble-based filters: Assimilation experiments with a shallow-water equation model. *J. Geophys. Res. Atmos.*, **113**, 1–12, doi:10.1029/2007JD008624.

Ying, Y., F. Zhang, Y. Ying, and F. Zhang, 2018: Potentials in Improving Predictability of Multiscale Tropical Weather Systems Evaluated through Ensemble Assimilation of Simulated Satellite-Based Observations. *J. Atmos. Sci.*, **75**, 1675–1698, doi:10.1175/JAS-D-17-0245.1.

Zapotocny, T. H., J. A. Jung, J. F. Le Marshall, R. E. Treadon, T. H. Zapotocny, J. A. Jung, J. F. Le Marshall, and R. E. Treadon, 2008: A Two-Season Impact Study of Four Satellite Data Types and Rawinsonde Data in the NCEP Global Data Assimilation System. *Wea. Forecasting*, **23**, 80–100, doi:10.1175/2007WAF2007010.1.

Zhang, F., Y. Weng, J. A. Sippel, Z. Meng, and C. H. Bishop, 2009: Cloud-resolving hurricane initialization and prediction through assimilation of doppler radar observations with an ensemble Kalman filter. *Mon. Wea. Rev.*, **137**, 2105–2125, doi:10.1175/2009MWR2645.1.

Zhang, M., and F. Zhang, 2012: E4DVar: Coupling an Ensemble Kalman Filter with Four-Dimensional Variational Data Assimilation in a Limited-Area Weather Prediction Model. *Mon. Wea. Rev.*, **140**, 587–600, doi:10.1175/MWR-D-11-00023.1.

——, ——, X.-Y. Huang, and X. Zhang, 2011: Intercomparison of an ensemble Kalman filter with three-and four-dimensional variational data assimilation methods in a limited-area model over the month of June 2003. *Mon. Wea. Rev.*, **139**, 566–572.

Zhao, Q., Q. Xu, Y. Jin, J. McLay, and C. Reynolds, 2015: Time-Expanded Sampling for Ensemble-Based Data Assimilation Applied to Conventional and Satellite Observations. *Wea. Forecasting*, **30**, 855–872, doi:10.1175/WAF-D-14-00108.1.

Zhou, L., S.-J. Lin, J.-H. Chen, L. M. Harris, X. Chen, and S. L. Rees, 2019: Toward Convective-Scale Prediction within the Next Generation Global Prediction System. *Bull. Amer. Meteor. Soc.*, **100**, 1225–1243, doi:10.1175/bams-d-17-0246.1.

Zhu, Y., J. Derber, A. Collard, D. Dee, R. Treadon, G. Gayno, and J. A. Jung, 2014: Enhanced radiance bias correction in the National Centers for Environmental Prediction's Gridpoint Statistical Interpolation data assimilation system. *Quart. J. Roy. Meteor. Soc.*, **140**, 1479–1492, doi:10.1002/qj.2233.

Zong, H., and L. Wu, 2015: Synoptic-scale influences on tropical cyclone formation within the western North Pacific monsoon trough. *Mon. Wea. Rev.*, **143**, 3421–3433.

Appendix A: Components of the VTSM- and VTSP-populated Background Ensemble Error Covariances

In the VTSM method, the background ensemble members valid at time $t-\tau$ and $t+\tau$, treated as the shifted background ensembles, are directly shifted to be valid at time t to supplement the original background ensemble at time t. If the original ensemble size is K, the VTSM method produces a populated background ensemble with size $3K$. The covariances from the VTSM-populated background ensemble $\mathbf{P}_{\mathrm{VTSM}}$ at time t can be derived as,

$$
\begin{aligned}
\mathbf{P}_{\mathrm{VTSM}} = \frac{1}{3K-1}\Big\{ & [\mathbf{X}_{t-\tau} - \mathrm{diag}(\overline{\mathbf{x}}_{\mathrm{VTSM}})\mathbf{1}_{n\times k} \quad \mathbf{X}_{t} - \mathrm{diag}(\overline{\mathbf{x}}_{\mathrm{VTSM}})\mathbf{1}_{n\times k} \quad \mathbf{X}_{t+\tau} - \mathrm{diag}(\overline{\mathbf{x}}_{\mathrm{VTSM}})\mathbf{1}_{n\times k}] \\
& [\mathbf{X}_{t-\tau} - \mathrm{diag}(\overline{\mathbf{x}}_{\mathrm{VTSM}})\mathbf{1}_{n\times k} \quad \mathbf{X}_{t} - \mathrm{diag}(\overline{\mathbf{x}}_{\mathrm{VTSM}})\mathbf{1}_{n\times k} \quad \mathbf{X}_{t+\tau} - \mathrm{diag}(\overline{\mathbf{x}}_{\mathrm{VTSM}})\mathbf{1}_{n\times k}]^{\mathrm{T}}\Big\} \\
= \frac{1}{3K-1}\Big\{ & [\mathbf{X}_{t-\tau} - \mathrm{diag}(\overline{\mathbf{x}}_{\mathrm{VTSM}})\mathbf{1}_{n\times k}][\mathbf{X}_{t-\tau} - \mathrm{diag}(\overline{\mathbf{x}}_{\mathrm{VTSM}})\mathbf{1}_{n\times k}]^{\mathrm{T}} \\
& + [\mathbf{X}_{t} - \mathrm{diag}(\overline{\mathbf{x}}_{\mathrm{VTSM}})\mathbf{1}_{n\times k}][\mathbf{X}_{t} - \mathrm{diag}(\overline{\mathbf{x}}_{\mathrm{VTSM}})\mathbf{1}_{n\times k}]^{\mathrm{T}} \\
& + [\mathbf{X}_{t+\tau} - \mathrm{diag}(\overline{\mathbf{x}}_{\mathrm{VTSM}})\mathbf{1}_{n\times k}][\mathbf{X}_{t+\tau} - \mathrm{diag}(\overline{\mathbf{x}}_{\mathrm{VTSM}})\mathbf{1}_{n\times k}]^{\mathrm{T}}\Big\} \\
= \frac{K-1}{3K-1}\Big\{ & \frac{1}{K-1}[\mathbf{X}_{t-\tau} - \mathrm{diag}(\overline{\mathbf{x}}_{t-\tau})\mathbf{1}_{n\times k}][\mathbf{X}_{t-\tau} - \mathrm{diag}(\overline{\mathbf{x}}_{t-\tau})\mathbf{1}_{n\times k}]^{\mathrm{T}} \\
& + \frac{1}{K-1}[\mathbf{X}_{t} - \mathrm{diag}(\overline{\mathbf{x}}_{t})\mathbf{1}_{n\times k}][\mathbf{X}_{t} - \mathrm{diag}(\overline{\mathbf{x}}_{t})\mathbf{1}_{n\times k}]^{\mathrm{T}} \\
& + \frac{1}{K-1}[\mathbf{X}_{t+\tau} - \mathrm{diag}(\overline{\mathbf{x}}_{t+\tau})\mathbf{1}_{n\times k}][\mathbf{X}_{t+\tau} - \mathrm{diag}(\overline{\mathbf{x}}_{t+\tau})\mathbf{1}_{n\times k}]^{\mathrm{T}}\Big\} \\
+ \frac{K}{3K-1}\Big\{ & [\overline{\mathbf{x}}_{t-\tau} - \overline{\mathbf{x}}_{\mathrm{VTSM}}][\overline{\mathbf{x}}_{t-\tau} - \overline{\mathbf{x}}_{\mathrm{VTSM}}]^{\mathrm{T}} \\
& + [\overline{\mathbf{x}}_{t} - \overline{\mathbf{x}}_{\mathrm{VTSM}}][\overline{\mathbf{x}}_{t} - \overline{\mathbf{x}}_{\mathrm{VTSM}}]^{\mathrm{T}} \\
& + [\overline{\mathbf{x}}_{t+\tau} - \overline{\mathbf{x}}_{\mathrm{VTSM}}][\overline{\mathbf{x}}_{t+\tau} - \overline{\mathbf{x}}_{\mathrm{VTSM}}]^{\mathrm{T}}\Big\},
\end{aligned}
\tag{A1}
$$

where, n is the dimension of the model state variables. $\mathbf{X}_{t-\tau}$, $\mathbf{X}_{t}$ and $\mathbf{X}_{t+\tau}$ are the background

ensemble matrices of $n\times K$ dimension at time t–τ, t and t+τ, respectively, and their corresponding background ensemble mean vector of $n\times 1$ dimension are denoted by $\overline{\mathbf{x}}_{t-\tau}$, $\overline{\mathbf{x}}_{t}$ and $\overline{\mathbf{x}}_{t+\tau}$. $\mathbf{1}_{n\times k}$ is a matrix with all the elements equal to 1 and its dimension denoted by the subscript (e.g., $n\times K$ dimension for the matrix $\mathbf{1}_{n\times k}$). $\overline{\mathbf{x}}_{\mathrm{VTSM}}=\frac{1}{3}(\overline{\mathbf{x}}_{t-\tau}+\overline{\mathbf{x}}_{t}+\overline{\mathbf{x}}_{t+\tau})$ is the VTSM-populated background ensemble mean equal to the average of the original background ensemble mean at the three different times. diag functions as converting a vector to a square diagonal matrix with the elements aligned on the diagonal. T is the matrix transpose sign. In Eq. (A1), the last step is derived by re-formulating each matrix-multiplication term in the second row of Eq. (A1). For example, the term corresponding to the time t is reformulated by,

$$
\begin{aligned}
&[\mathbf{X}_t-\mathrm{diag}(\overline{\mathbf{x}}_{\mathrm{VTSM}})\mathbf{1}_{n\times k}][\mathbf{X}_t-\mathrm{diag}(\overline{\mathbf{x}}_{\mathrm{VTSM}})\mathbf{1}_{n\times k}]^{\mathrm{T}}\\
&\quad=\mathbf{X}_t\mathbf{X}_t^{\mathrm{T}}-\mathbf{X}_t\mathbf{1}_{k\times n}\mathrm{diag}(\overline{\mathbf{x}}_{\mathrm{VTSM}})-\mathrm{diag}(\overline{\mathbf{x}}_{\mathrm{VTSM}})\mathbf{1}_{n\times k}\mathbf{X}_t^{\mathrm{T}}+\mathrm{diag}(\overline{\mathbf{x}}_{\mathrm{VTSM}})\mathbf{1}_{n\times k}\mathbf{1}_{k\times n}\mathrm{diag}(\overline{\mathbf{x}}_{\mathrm{VTSM}})\\
&\quad=\mathbf{X}_t\mathbf{X}_t^{\mathrm{T}}-\mathbf{X}_t\mathbf{1}_{k\times n}\mathrm{diag}(\overline{\mathbf{x}}_{t})-\mathrm{diag}(\overline{\mathbf{x}}_{t})\mathbf{1}_{n\times k}\mathbf{X}_t^{\mathrm{T}}+\mathrm{diag}(\overline{\mathbf{x}}_{t})\mathbf{1}_{n\times k}\mathbf{1}_{k\times n}\mathrm{diag}(\overline{\mathbf{x}}_{t})\\
&\qquad+\mathbf{X}_t\mathbf{1}_{k\times n}\mathrm{diag}(\overline{\mathbf{x}}_{t})+\mathrm{diag}(\overline{\mathbf{x}}_{t})\mathbf{1}_{n\times k}\mathbf{X}_t^{\mathrm{T}}-\mathrm{diag}(\overline{\mathbf{x}}_{t})\mathbf{1}_{n\times k}\mathbf{1}_{k\times n}\mathrm{diag}(\overline{\mathbf{x}}_{t})\\
&\qquad-\mathbf{X}_t\mathbf{1}_{k\times n}\mathrm{diag}(\overline{\mathbf{x}}_{\mathrm{VTSM}})-\mathrm{diag}(\overline{\mathbf{x}}_{\mathrm{VTSM}})\mathbf{1}_{n\times k}\mathbf{X}_t^{\mathrm{T}}+\mathrm{diag}(\overline{\mathbf{x}}_{\mathrm{VTSM}})\mathbf{1}_{n\times k}\mathbf{1}_{k\times n}\mathrm{diag}(\overline{\mathbf{x}}_{\mathrm{VTSM}})\\
&\quad=[\mathbf{X}_t-\mathrm{diag}(\overline{\mathbf{x}}_{t})\mathbf{1}_{n\times k}][\mathbf{X}_t-\mathrm{diag}(\overline{\mathbf{x}}_{t})\mathbf{1}_{n\times k}]^{\mathrm{T}}\\
&\qquad+K\ \mathrm{diag}(\overline{\mathbf{x}}_{t})\mathbf{1}_{n\times n}\mathrm{diag}(\overline{\mathbf{x}}_{t})+K\ \mathrm{diag}(\overline{\mathbf{x}}_{t})\mathbf{1}_{n\times n}\mathrm{diag}(\overline{\mathbf{x}}_{t})-K\ \mathrm{diag}(\overline{\mathbf{x}}_{t})\mathbf{1}_{n\times n}\mathrm{diag}(\overline{\mathbf{x}}_{t})\\
&\qquad-K\ \mathrm{diag}(\overline{\mathbf{x}}_{t})\mathbf{1}_{n\times n}\mathrm{diag}(\overline{\mathbf{x}}_{\mathrm{VTSM}})-K\ \mathrm{diag}(\overline{\mathbf{x}}_{\mathrm{VTSM}})\mathbf{1}_{n\times n}\mathrm{diag}(\overline{\mathbf{x}}_{t})+K\ \mathrm{diag}(\overline{\mathbf{x}}_{\mathrm{VTSM}})\mathbf{1}_{n\times n}\mathrm{diag}(\overline{\mathbf{x}}_{\mathrm{VTSM}})\\
&\quad=[\mathbf{X}_t-\mathrm{diag}(\overline{\mathbf{x}}_{t})\mathbf{1}_{n\times k}][\mathbf{X}_t-\mathrm{diag}(\overline{\mathbf{x}}_{t})\mathbf{1}_{n\times k}]^{\mathrm{T}}\\
&\qquad+K\ \mathrm{diag}(\overline{\mathbf{x}}_{t})\mathbf{1}_{n\times n}\mathrm{diag}(\overline{\mathbf{x}}_{t}-\overline{\mathbf{x}}_{\mathrm{VTSM}})+K\ \mathrm{diag}(\overline{\mathbf{x}}_{\mathrm{VTSM}})\mathbf{1}_{n\times n}\mathrm{diag}(\overline{\mathbf{x}}_{\mathrm{VTSM}}-\overline{\mathbf{x}}_{t})\\
&\quad=[\mathbf{X}_t-\mathrm{diag}(\overline{\mathbf{x}}_{t})\mathbf{1}_{n\times k}][\mathbf{X}_t-\mathrm{diag}(\overline{\mathbf{x}}_{t})\mathbf{1}_{n\times k}]^{\mathrm{T}}+K\ \mathrm{diag}(\overline{\mathbf{x}}_{t}-\overline{\mathbf{x}}_{\mathrm{VTSM}})\mathbf{1}_{n\times n}\mathrm{diag}(\overline{\mathbf{x}}_{t}-\overline{\mathbf{x}}_{\mathrm{VTSM}})\\
&\quad=[\mathbf{X}_t-\mathrm{diag}(\overline{\mathbf{x}}_{t})\mathbf{1}_{n\times k}][\mathbf{X}_t-\mathrm{diag}(\overline{\mathbf{x}}_{t})\mathbf{1}_{n\times k}]^{\mathrm{T}}+K\ [\overline{\mathbf{x}}_{t}-\overline{\mathbf{x}}_{\mathrm{VTSM}}][\overline{\mathbf{x}}_{t}-\overline{\mathbf{x}}_{\mathrm{VTSM}}]^{\mathrm{T}}
\end{aligned}
\tag{A2}
$$

Eq. (A1) shows that if the original ensemble size is large enough, the VTSM-populated background ensemble covariances can be obtained by summing up two components. One is an

approximate average of the original background ensemble covariances at time $t-\tau$, t and $t+\tau$ (the terms within the first braces in the last row of Eq. (A1)). The second component is approximately equal to averaging three matrices, respectively, obtained by an outer product of the vector representing the difference between the VTSM-populated ensemble mean and the original ensemble mean at the three different times. Since the VTSM-populated background ensemble mean is calculated as the average of the original background ensemble means at three different times, the second component can also be interpreted as the contribution from the background ensemble mean differences between the original and shifted background ensembles (the terms within the second braces in the last row of Eq. (A1)).

In the VTSP method, the populated background ensemble perturbations at time t are constructed by shifting the original background ensemble perturbations valid at time $t-\tau$ and $t+\tau$ to the time t. Correspondingly, the covariances from the VTSP-populated background ensembles with size of $3K$ at time t can be expressed as[2],

[2] In the calculations of the VTSP-populated background ensemble covariances of Eq. (A3), a factor of ($3K$-3) is supposed to be used as the denominator in Eq. (A3) as explained in the texts. In the practical implementation of the VTSP method, a factor of ($3K$-1) was used as the denominator in Eq. (A3). But the misuse of the factor of ($3K$-1) only causes a very tiny error by a factor less than 1%.

$$\mathbf{P}_{\mathrm{VTSP}} = \frac{1}{3K-3}\Big\{[\mathbf{X}_{t-\tau} - \mathrm{diag}(\overline{\mathbf{x}}_{t-\tau})\mathbf{1}_{n\times k} \quad \mathbf{X}_{t} - \mathrm{diag}(\overline{\mathbf{x}}_{t})\mathbf{1}_{n\times k} \quad \mathbf{X}_{t+\tau} - \mathrm{diag}(\overline{\mathbf{x}}_{t+\tau})\mathbf{1}_{n\times k}]$$
$$[\mathbf{X}_{t-\tau} - \mathrm{diag}(\overline{\mathbf{x}}_{t-\tau})\mathbf{1}_{n\times k} \quad \mathbf{X}_{t} - \mathrm{diag}(\overline{\mathbf{x}}_{t})\mathbf{1}_{n\times k} \quad \mathbf{X}_{t+\tau} - \mathrm{diag}(\overline{\mathbf{x}}_{t+\tau})\mathbf{1}_{n\times k}]^{\mathrm{T}}\Big\}$$
$$= \frac{1}{3}\Big\{\frac{1}{K-1}[\mathbf{X}_{t-\tau} - \mathrm{diag}(\overline{\mathbf{x}}_{t-\tau})\mathbf{1}_{n\times k}][\mathbf{X}_{t-\tau} - \mathrm{diag}(\overline{\mathbf{x}}_{t-\tau})\mathbf{1}_{n\times k}]^{\mathrm{T}}$$
$$+ \frac{1}{K-1}[\mathbf{X}_{t} - \mathrm{diag}(\overline{\mathbf{x}}_{t})\mathbf{1}_{n\times k}][\mathbf{X}_{t} - \mathrm{diag}(\overline{\mathbf{x}}_{t})\mathbf{1}_{n\times k}]^{\mathrm{T}}$$
$$+ \frac{1}{K-1}[\mathbf{X}_{t+\tau} - \mathrm{diag}(\overline{\mathbf{x}}_{t+\tau})\mathbf{1}_{n\times k}][\mathbf{X}_{t+\tau} - \mathrm{diag}(\overline{\mathbf{x}}_{t+\tau})\mathbf{1}_{n\times k}]^{\mathrm{T}}\Big\}. \tag{A3}$$

In Eq. (A3), because the three groups of the original background ensemble perturbations are calculated from its own background ensemble means, 3 degrees of freedom are removed from the VTSP-populated background ensemble. So $(3K-3)$ instead of $(3K-1)$ is supposed to be used as the denominator in Eq. (A3) to obtain the best unbiased covariances of the population from VTSP-populated background ensemble sample. It shows the VTSP-populated background ensemble error covariances are equal to the average of the original background ensemble covariances at the three different times.

As noted, the calculations of the VTSM- and VTSP-populated background ensemble covariances in Eqs (A1) and (A3) can be also applied to the scenarios of more than three time levels.

Appendix B: Illustration of collecting nearly independent samples for paired permutation test

In the global forecast verification in sections 5.4.3, 5.4.4 and 5.5.1, at a particular forecast lead time and model pressure level, the nearly independent forecast error difference samples were collected to perform the paired permutation test following Wang and Bishop (2005).

(1) The forecast error time series over a total of 96 data assimilation cycles in a particular experiment were first obtained by verifying against the rawinsonde observations or ECMWF reanalyses over the full global domain.

(2) The full global domain was divided into *m* subdomains. The forecast error time series were averaged over each of *m* subdomains to obtain *m* sets of forecast error time series in a particular experiment. *m* sets of forecast error difference time series were collected by comparing two experiments such as W1-1000 and W1-Ope.

(5) Correlations among *m* sets of forecast error difference time series were calculated to determine the number of *m* so that they had a weak correlation (e.g., <=0.3). As a result, it produced *m* sets of spatially nearly independent forecast error difference time series.

(6) The lagged autocorrelations were calculated for each of *m* sets of spatially nearly independent forecast error difference time series. A lag length was further determined using a weak autocorrelation (e.g., <=0.3). Then *m* sets of spatially and temporally nearly independent forecast error difference time series were collected from each of *m* sets of spatially nearly independent forecast error difference time series by skipping the samples by a number equal to the lag length.

(7) Finally, *m* sets of spatially and temporally nearly independent forecast error

difference time series were pooled together to perform the paired permutation tests.

In the six-hour forecast verification against the rawinsodnes in section 5.4.3, a large number of m=595 was selected to compensate the reduction of effective sample size due to the non-uniform rawinsonde observation coverage over the globe and at different model pressure levels. Figure B.1a shows an example of correlations among m=595 sets of six-hour temperature forecast error difference time series at 500 hPa between the W1-1000 and W1-Ope. They generally show relatively weak correlations. Similar results were also found at other model pressure levels and for wind forecasts. Figure B.1b shows the lagged autocorrelation of m=595 sets of six-hour temperature forecast error difference time series as in Fig. B.1a. Their obtained lag length varies for each set of samples. In the 0-5-day global forecast verification against the ECMWF reanalyses in sections 5.4.3 and 5.5.1, m=162 was selected. Figure B.2a,b shows an example of correlations and lagged autocorrelations, respectively, among m=161 sets of one-day temperature forecast error difference time series at 500 hPa between the W1-1000 and W1-Ope. They generally show relatively weak correlations. The obtained lag length varies for each set of samples. Similar results were also found at other pressure levels and wind forecasts.

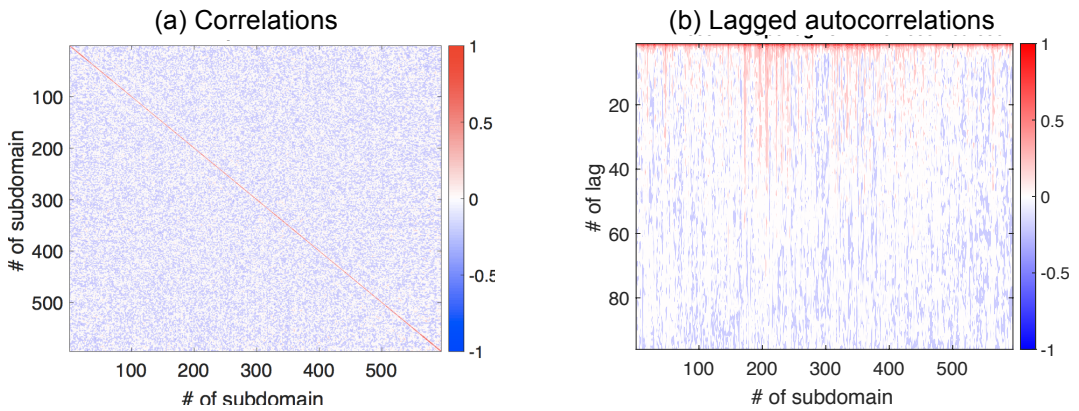

Figure B.1 (a) Correlations and (b) lagged autocorrelations among m=595 sets of six-hour temperature forecast error difference time series at 500 hPa between the W1-1000 and W1-Ope. Each set of samples were collected from the corresponding subdomain. The lag number in (b) is denoted on the Y-axis.

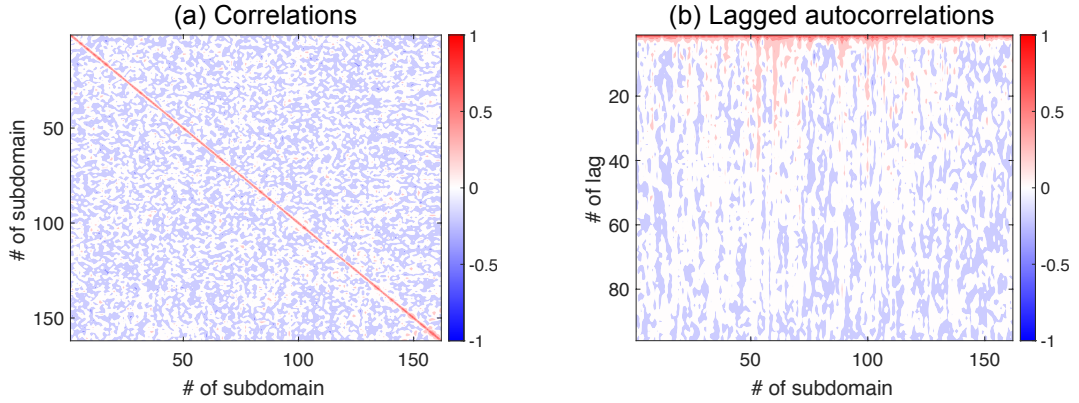

Figure B.2 (a) Correlations and (b) lagged autocorrelations among m=162 sets of one-day temperature forecast error difference time series at 500 hPa between the W1-1000 and W1-Ope. Each set of samples were collected from the corresponding subdomain. The lag number in (b) is denoted on the Y-axis.

In the tropical storm track forecasts, each storm was treated as an independent case. The lagged autocorrelations were further calculated to determine a lag length in each set of track error difference time series between two experiments. Since there were a limited number of tropical storms and they did not last over the full one-month cycling period, the original tropical

storm sample size is much smaller than those in the global forecast evaluation. To remedy this limitation and collect more samples, a correlation of 0.7 was used to determine the lag length here. Because the tropical storms lasted for different period, it resulted in track error difference time series with different length. This makes it less feasible to plot correlations as in the global forecast verification. So plots of correlations and lagged autocorrelations in the tropical storm evaluation are not shown here.

In the power spectral evaluation in section 5.6.3, it produced error total energy difference time series between two experiments at a particular model level, total wavenumber and forecast lead time. In this evaluation, each model level was treated as spatially independent. Figures B.3e,f show the correlations and lagged autocorrelations of m=16 (i.e., the number of model levels) sets of error total energy difference time series at total wavenumber ten and at one-day forecast lead time. They show relatively small correlations. The obtained lag length also varies for each set of samples. Similar results were also found at other total wavenumbers and forecast lead times.

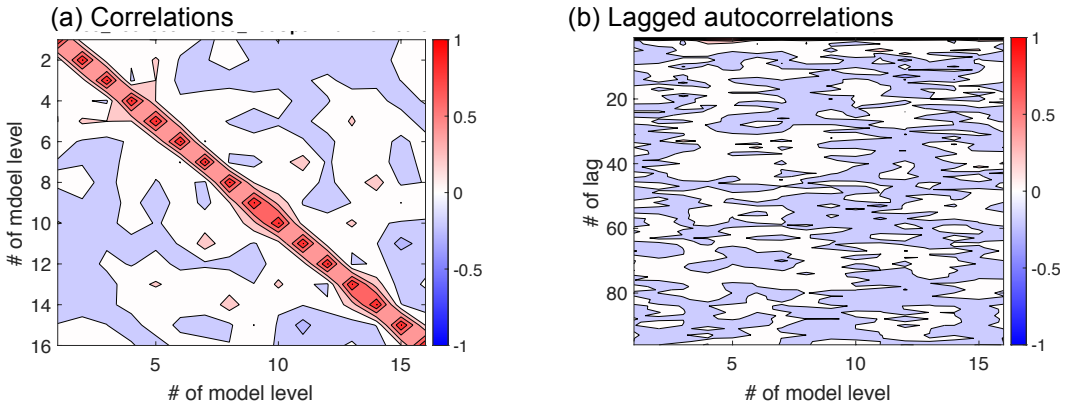

Figure B.3 (a) Correlations and lagged autocorrelations of m=16 sets of error total energy difference time series between W1-1000 and W1-Ope at total wavenumber ten and at one-day forecast lead time. Each set of samples were collected from the corresponding model level. The lag number in (b) is denoted on the Y-axis.