ABSTRACT: A hybrid three-dimensional ensemble–variational (En3D-Var) data assimilation system has been developed to explore incorporating information from an 11-member regional ensemble prediction system, which is dynamically downscaled from a global ensemble system, into a 3-hourly cycling convective-scale data assimilation system over the western Maritime Continent. From the ensemble, there exists small-scale ensemble perturbation structures associated with positional differences of tropical convection, but these structures are well represented only after the downscaled ensemble forecast has evolved for at least 6 h due to spinup. There was also a robust moderate negative correlation between total specific humidity and potential temperature background errors, presumably because of incorrect vertical motion in the presence of clouds. Time shifting of the ensemble perturbations, by using those available from adjacent cycles, helped to ameliorate the sampling error prevalent in their raw autocovariances. Monthlong hybrid En3D-Var trials were conducted using different weights assigned to the ensemble-derived and climatological background error covariances. The forecast fits to radiosonde relative humidity and wind observations were generally improved with hybrid En3D-Var, but in all experiments, the fits to surface observations were degraded compared to the baseline 3D-Var configuration. Over the Singapore radar domain, there was a general improvement in the precipitation forecasts, especially when the weighting toward the climatological background error covariance was larger, and with the additional application of time-shifted ensemble perturbations. Future work involves consolidating the ensemble prediction and deterministic system, by centering the ensemble prediction system on the hybrid analysis, to better represent the analysis and forecast uncertainties.

KEYWORDS: Maritime Continent; Ensembles; Numerical weather prediction/forecasting; Data assimilation

1. Introduction

At the Meteorological Service Singapore (MSS), a hybrid ensemble–variational data assimilation system has been developed to explore incorporating information from an ensemble prediction system (SINGV-EPS; Porson et al. 2019) into a variational data assimilation system (SINGV-DA; Heng et al. 2020). Such hybrid ensemble–variational methods have recently gained traction, used in both global numerical weather prediction (NWP) systems (Buehner et al. 2013; Clayton et al. 2013; Kuhl et al. 2013; Wang et al. 2013; Bonavita et al. 2016; Kadowaki et al. 2020) and regional NWP systems (Zhang and Zhang 2012; Gustafsson et al. 2014; Ito et al. 2016; Montmerle et al. 2018; Caron et al. 2019; Bédard et al. 2020) at leading operational NWP centers. Different centers employ their own flavor of hybrid ensemble–variational methods, due to many possible permutations in the design of the ensemble prediction system and/or variational data assimilation system. Bannister (2017) provides a thorough overview of hybrid ensemble–variational methods used in operational systems.

Apart from MSS, there are only a handful of research and operational centers that maintain a convective-scale NWP system over the western Maritime Continent\(^1\) (Centre for Climate Research Singapore 2019). Very few centers have incorporated data assimilation, and these centers typically apply traditional variational methods partly due to the lack of a suitable high-resolution ensemble which is needed for the application of hybrid ensemble–variational methods. In traditional variational data assimilation, the characterization of the background errors often relies on the assumption of ergodicity by using climatological error statistics. Modeling the climatological background error covariance matrix also requires further assumptions of homogeneity, isotropy, and balance constraints in the background error statistics (Bannister 2008). These are necessary to prescribe a parameterized model of the climatological background error covariance matrix so that the variational data assimilation problem becomes computationally feasible. Over the western Maritime Continent, these assumptions may be often violated by the presence of nonlinear convective processes, land–sea interactions and local orographic effects (Lee and Huang 2022). On the other hand, the characterization of the ensemble-derived background error covariance matrix, usually within an ensemble Kalman filter (EnKF) framework (Evensen 1994),

Publisher’s Note: This article was revised on 3 May 2023 correct a typographical error in section 4e.

\(^1\)The term Maritime Continent is a nickname for the Indo-Pacific archipelago, encompassing many islands, peninsulas, and the surrounding seas of Southeast Asia.
does not require these assumptions as the background error statistics can be estimated directly from the model states. The error statistics may also contain meaningful flow-dependent error structures related to the short-lived tropical weather phenomena over the region. However, since the degrees of freedom for the model state is usually far greater than the ensemble size used to estimate the error statistics (resulting in a rank-deficient matrix), sampling noise and spurious long-range correlations may be present, particularly for smaller ensembles. To address sampling noise, Houtekamer and Mitchell (2001) proposed computing the Schur product of a localization matrix (a correlation matrix) and the ensemble-derived background error covariance matrix. This effectively increases the rank of the ensemble-derived background error covariance matrix. However, the use of localization can unintentionally introduce dynamical imbalances into the system (Lorenc 2003), which may be detrimental.

The use of a weighted combination of both estimates of background error statistics in a hybrid ensemble–variational data assimilation framework was proposed by Hamill and Snyder (2000). The main idea is to alleviate the limitations and maximize the advantages offered by individual ensemble-based or variational methods themselves. The ensemble-derived background error covariance matrix can augment the climatological background error covariance matrix with its flow-dependent error structures, while the climatological background error covariance matrix can ameliorate the sampling noise issues associated with the ensemble-derived background error covariance matrix. Previous studies have provided evidence that the weighted combination results in improved verification scores over individual ensemble-based or variational methods. The optimal weighting combination varies between studies because of a multitude of factors including ensemble size (Hamill and Snyder 2000), localization space and length scales (Montmerle et al. 2018; Caron et al. 2019), design of the ensemble (e.g., EnKF or other deterministic analysis ensemble methods; Tippett et al. 2003), and variational method applied (e.g., three- or four-dimensional; 3D-Var or 4D-Var). Most operational centers rely on empirical tuning to find the optimal weighting.

In this article, we describe the development of a hybrid ensemble–variational data assimilation system over the western Maritime Continent, which is—to our knowledge—the first of its kind over this region. The initial implementation of SINGV-EPS is a simple 11-member ensemble prediction system dynamically downscaled from preselected European Centre for Medium-Range Weather Forecasts (ECMWF) global ensemble members every 12 h over the western Maritime Continent. SINGV-DA is a 3-hourly cycling three-dimensional first guess at appropriate time variational system (3D-Var FGAT) over the same domain. There is also no centering of SINGV-EPS on the SINGV-DA analyses, so the information flow is one-way (i.e., from SINGV-EPS to SINGV-DA during the estimation of the ensemble-derived background error statistics). Section 2 describes the hybrid ensemble–variational formulation and implementation at MSS, along with more details on SINGV-EPS and SINGV-DA. We explore the structures of the ensemble-derived background error statistics and discuss their relevance in section 3. Section 4 contains a description of the monthlong trials, which are conducted to seek a suitable configuration for operational implementation. We also discuss the justification of the tuning parameters for the western Maritime Continent context and include verification scores from the empirical tuning of the hybrid three-dimensional ensemble–variational (En3D-Var) system (see section 2c for nomenclature) at MSS.

2. Hybrid En3D-Var formulation
a. Traditional 3D-Var FGAT in SINGV-DA

SINGV-DA is a convective-scale regional NWP data assimilation system that uses a horizontal grid spacing of approximately 1.5 km, with 80 vertical levels up to 38.5 km over the western Maritime Continent (see Fig. 2 for the domain). The lateral boundary conditions (LBCs) are provided by ECMWF analyses and forecasts every 6 h (0000, 0600, 1200, and 1800 UTC). SINGV-DA is based on the Met Office (UKMO) Unified Model framework (Tang et al. 2013), and is designed as a 3-hourly cycling 3D-Var FGAT system. Assimilated observations are retrieved from WMO’s Global Telecommunication System. These include radiosondes, surface and aircraft observations, all-sky and clear-sky satellite radiance observations, satellite-derived wind observations, and satellite-derived pseudo–cloud observations (see Heng et al. 2020 for the full list). The observation error profiles are retrieved from the UKMO.

At each SINGV-DA cycle, we seek an “optimal” state \( \mathbf{x}^o \) that minimizes a cost function \( J(\mathbf{x}) \) (e.g., Kalnay 2003). The cost function is usually nonquadratic because of the often nonlinear forecast model \( \mathbf{M} \) and nonlinear observation operator \( \mathbf{H} \), which appear in one of its components: the departure of the “optimal” state with respect to observations; observation penalty. Like most variational systems, SINGV-DA implements an incremental formulation of the cost function (Courtier et al. 1994), which requires a linearization of \( \mathbf{M} \) and \( \mathbf{H} \) around a reference state \( \mathbf{x}^t \) and formulating the problems in terms of increments \( \delta \mathbf{x} \) to \( \mathbf{x}^t \) in a series of outer loops. In SINGV-DA, we do not include an imbalance penalty (e.g., as included in Clayton et al. 2013; Milan et al. 2020) and assume a perfect model. Since FGAT is used, the linearized forecast model \( \mathbf{M} = \mathbf{I} \), an identity matrix. The (strong-constraint) incremental form of the 3D-Var FGAT cost function is thus given by

\[
J(\delta \mathbf{x}) = J_b + J_o, \\
= \frac{1}{2}(\delta \mathbf{x} - \delta \mathbf{x}^t)^T \mathbf{B}^{-1}(\delta \mathbf{x} - \delta \mathbf{x}^t) \\
+ \frac{1}{2}(\delta \mathbf{x} - \delta \mathbf{x}^t)^T \mathbf{R}^{-1}(\delta \mathbf{x} - \delta \mathbf{x}^t) - \mathbf{H}^T \mathbf{R}^{-1} \mathbf{H} \delta \mathbf{x} - \mathbf{H} \delta \mathbf{x}^t, \\
(1)
\]

where \( J_b \) and \( J_o \) are the background and observation penalties, respectively; \( \delta \mathbf{x}^t \) is the difference between the background and \( \mathbf{x}^t \); \( \mathbf{B} \) and \( \mathbf{R} \) are the background and observation error covariance matrices; and \( \mathbf{H} \) is the linearized observation operator. The innovation \( \mathbf{d} \) is defined as

\[
\mathbf{d} = \mathbf{y}^o - \mathbf{H}[\mathbf{M}(\mathbf{x}^t)],
\]

(2)

where \( \mathbf{y}^o \) is the observations vector. As SINGV-DA uses the observations processing and data assimilation framework
from the UKMO, the computation of the observation penalty is the same as in Eq. (2) of Clayton et al. (2013), except that the linear perturbation forecast model (analogous to M) is not required in 3D-Var FGAT (hereafter referred to as 3D-Var for brevity). These details are also discussed in Lorenc et al. (2000).

The conditioning of the cost function minimization [Eq. (1)] can be improved by introducing a control variable transform \( \mathbf{U} \) (a “square root” of \( \mathbf{B} \); i.e., \( \mathbf{B} = \mathbf{UU}^T \)). This avoids the need to compute \( \mathbf{B}^{-1} \). The increment \( \delta \mathbf{x} \) can be expressed as

\[
\delta \mathbf{x} = \mathbf{U} \delta \mathbf{X}.
\]

where \( \delta \mathbf{X} \) is the control vector. SINGV-DA adopts the same control variable transform (for full transforms, see Lorenc et al. 2000) as the UKMO limited-area model (LAM; UKV; Tang et al. 2013) prior to July 2017. The control variables used in \( \mathbf{U} \) are streamfunction, velocity potential, unbalanced pressure and a nonlinear transformed humidity (Ingleby et al. 2013). Linear balance is used as a balance constraint, so for the western Maritime Continent the geostrophically balanced pressure is small and pressure is near-wholly unbalanced (i.e., horizontal wind and pressure background errors are assumed to be almost uncorrelated). The same vertical modes in the vertical transform are used for each horizontal mode in the horizontal transform, so \( \mathbf{B} \) is homogeneous in the domain. Note that the additional vertical adaptive grid transform (e.g., used in Milan et al. 2020) is not applied. The same \( \mathbf{B} \) (strictly speaking the same \( \mathbf{U} \)) is used in the variational minimization for each cycle, thus \( \mathbf{B} \) is often referred to as the climatological background covariance matrix \( \mathbf{B}_c \). Time stationarity of \( \mathbf{B}_c \) is assumed even though the true forecast errors of the system (and their probability distribution function) may vary according to the flow conditions associated with various tropical weather phenomena.

The training data for calibrating \( \mathbf{B}_c \) for SINGV-DA was generated using the lagged National Meteorological Center (NMC) method (Široká et al. 2003), which is an extension of the original NMC method (Parrish and Derber 1992). The main difference is that the lagged NMC method excludes error sources from the driving model in regional NWP, by ensuring that the forecast differences along the boundaries are zero. The same set of LBCs are used in forecast pairs that are valid at the same time. This method is suitable for SINGV-DA as it retains the mesoscale and small-scale background error structures and does not reanalyze the scales already treated by the global model (Široká et al. 2003). Other studies with LAMs have also opted for the lagged NMC method (Sadiki and Fischer 2005; Montmerrer et al. 2006; Stanesic et al. 2019). The training data were based on the differences between 12- and 6-h forecasts valid at the same time, over the period 22 January–6 March 2013. However, while SINGV-DA uses a 1.5-km horizontal grid spacing, the training data were generated using a preliminary forecast model that uses a 4.5-km grid spacing instead. This discrepancy is further discussed in section 4e.

The characteristics of \( \mathbf{B}_c \) that is used in SINGV-DA have been explored in Heng et al. (2020) and Lee and Huang (2020). We remark that the variances in \( \mathbf{B}_c \) are simply the in-sample variances from the training data of lagged forecast differences. However, since the inception of SINGV-DA, the original \( \mathbf{B}_c \) has not been replaced due to poorer verification scores with other calibrated covariances. Heng et al. (2020) also describes other aspects of SINGV-DA, including the observational coverage, the application of an incremental analysis update (IAU) scheme (Bloom et al. 1996), satellite bias corrections, and other relevant SINGV-DA details.

b. Ensemble-derived background error covariances from SINGV-EPS

SINGV-EPS is a convective-scale 11-member ensemble prediction system that uses a horizontal grid spacing of 4.5 km. Each ensemble member is dynamically downscaled over the western Maritime Continent (near-identical domain and with the same model settings in each ensemble member as SINGV-DA, but at a lower resolution) using the corresponding global ECMWF ensemble member analysis and forecast. As ECMWF offers 51 global ensemble members (at ~18-km grid spacing; see Buizza and Richardson 2017) and SINGV-EPS only requires 11, the initial conditions and LBCs for SINGV-EPS are retrieved from the same members (simple fixed preselection, based on the first 11 odd number indexes of the 51 members) for reinitialization every 12 h. Since SINGV-EPS does not incorporate data assimilation, it is uninformed of the SINGV-DA observation network, unlike in other ensembles within the EnKF framework that can account for observation uncertainty and network. Only the nature of the dynamical error growth about the ensemble mean is represented during the estimation of the ensemble-derived background error covariance matrix. This is a limitation of the initial implementation, and future work on SINGV-EPS should address it.

An ensemble prediction system like SINGV-EPS allows for the representation of flow-dependent forecast errors due to varying flow conditions, which can be estimated using the ensemble forecast trajectories. Only the necessary ensemble forecast fields (zonal and meridional wind, potential temperature, a density term, pressure, total specific humidity) are required for reconfiguration onto the 3D-Var assimilation grid, usually at a coarser resolution and smaller than the forecast domain. One may compute a rectangular matrix \( \mathbf{X}' \) whose columns contain the scaled differences between the ensemble forecasts and the ensemble mean:

\[
\mathbf{X}' = \frac{1}{\sqrt{N-1}}(x_1^f - x_1^m, x_2^f - x_2^m, \ldots, x_N^f - x_N^m) = (x_1'^f, x_2'^f, \ldots, x_N'^f),
\]

where \( N \) is the number of ensemble members, \( x_i^f \) is the \( i \)th member forecast and \( x_i^m \) is ensemble mean valid at time \( t \) and \( x_i'^f \) is the \( k \)th scaled ensemble perturbation. The discrete validity time of the ensemble perturbations are chosen to correspond to the relevant cycle times (eight cycles a day) in SINGV-DA. Since SINGV-EPS is not coupled to SINGV-DA and does not require the deterministic analysis (no centering) to reinitialize the ensemble, these ensemble perturbations can be generated prior to running SINGV-DA.
The raw ensemble-derived background error covariance $\mathbf{P}_f^e$ is explicitly given by the outer product:

$$
\mathbf{P}_f^e = \mathbf{X}_f^e \mathbf{X}_f^e^T.
$$

(5)

The number of ensemble members only has $N$ degrees of freedom to fit the observations, so $\mathbf{P}_f^e$ is rank-deficient and is contaminated by sampling noise. Typically, a Schur product (Houtekamer and Mitchell 2001) with a localization matrix (L) is used to damp any possible spurious long-range correlations:

$$
\mathbf{B}_c = \mathbf{L} \cdot \mathbf{P}_f^e,
$$

(6)

where $\mathbf{B}_c$ is the ensemble-derived background error covariance matrix after localization and the $\circ$ operator denotes the Schur product (or Hadamard product), which conducts an element-wise product of two same-sized matrices. The design of $\mathbf{L}$ is further discussed in section 2d.

c. Hybrid background error covariance

The hybrid background error covariance matrix $\mathbf{B}_h$ is a linear combination of $\mathbf{B}_c$ and $\mathbf{B}_e$ (Hamill and Snyder 2000), in the following form:

$$
\mathbf{B}_h = \beta_1^e \mathbf{B}_e + \beta_2^e \mathbf{B}_c,
$$

(7)

where $\beta_1^e$ and $\beta_2^e$ are (positive) scalar weights that are usually determined empirically. These weights are often chosen to sum to unity, although it is not mandatory. It is not feasible to explicitly compute $\mathbf{B}_h$ from individual components for an NWP system like SINGV-DA, so the alpha control variable approach of Lorenc (2003) is employed which constructs an implied version of Eq. (6) using a modified version of Eq. (1). Wang et al. (2007) demonstrates how both approaches yield equivalent results.

The modified cost function [extension of Eq. (1)] is given by

$$
J(\delta \mathbf{x}, \alpha_1, \alpha_2, \ldots, \alpha_N) = J_b + J_o + J_c
= \frac{1}{2} (\delta \mathbf{x} - \delta \mathbf{x}^o)^T \mathbf{B}_c^{-1} (\delta \mathbf{x} - \delta \mathbf{x}^o)
+ \frac{1}{2} (\mathbf{H} \delta \mathbf{x} - \mathbf{d})^T \mathbf{R}^{-1} (\mathbf{H} \delta \mathbf{x} - \mathbf{d})
+ \frac{1}{2} \sum_{k=1}^N \alpha_k^e \mathbf{L}^{-1} \alpha_k^e,
$$

(8)

where $J_c$ is the ensemble penalty, and $\alpha_k$ is an alpha field with dimensions the same as the state size, associated with the $k$th ensemble member. Correspondingly, a modification of Eq. (3) is also required to include the ensemble contribution from a linear combination of ensemble perturbations:

$$
\delta \mathbf{x} = \beta_c \mathbf{U} \delta \mathbf{x} + \beta_c \sum_{k=1}^N \alpha_k^e \mathbf{x}_f^e.
$$

(9)

The alpha fields essentially control the contribution of each ensemble perturbation to the analysis increment. Like $\mathbf{B}_e$, $\mathbf{L}$ can be partitioned into a “square-root” matrix $\mathbf{U}^o$ (i.e., $\mathbf{L} = \mathbf{U}^o \mathbf{U}^o^T$), the alpha control variable transform, which can be applied to an alpha control vector associated with the $k$th ensemble member ($\chi_f^e$), so Eq. (9) becomes

$$
\mathbf{\delta} \mathbf{x} = \beta_c \mathbf{U} \mathbf{\delta} \mathbf{x} + \beta_c \sum_{k=1}^N \chi_f^e \circ (\mathbf{U}^o \chi_f^e).
$$

(10)

Note that using $\mathbf{U}^o$ avoids the need to compute $\mathbf{L}^{-1}$ in Eq. (8). Following the naming convention of Bannister (2017), this approach is termed as hybrid En3D-Var. Note that this is simply 3D-Var FGAT using $\mathbf{B}_h$, or hybrid 3D-Var-Ben following the nomenclature of Lorenc (2013).

Since SINGV-DA is a 3-hourly cycling system, $\mathbf{B}_h$ is required at each cycle so there needs to be valid ensemble perturbations from SINGV-EPS every 3 h. However, SINGV-EPS is initialized once every 12 h, so the ensemble forecast trajectories downscaled from a given set of ECMWF driving data have to provide ensemble perturbations for multiple SINGV-DA cycles that fall within the 12-h window until the next set of ECMWF driving data is available. Figure 1 shows the schematic diagram illustrating the design of the hybrid En3D-Var system and flow of information from SINGV-EPS to SINGV-DA. As an example, the 12-h ensemble forecasts are used to compute the ensemble perturbations [Eq. (4)] for the 0000 and 1200 UTC SINGV-DA cycles, and the 9-h forecasts are used for the 0900 and 2100 UTC SINGV-DA cycles. Evidently, a key limitation is that the ensemble statistics are calculated from a longer forecast range (for 6–12-h forecasts) than required for a 3-hourly cycling SINGV-DA system requiring 3-h forecast error statistics. Since SINGV-EPS is also downscaled from global driving data, spinup may also affect the ensemble statistics so the variances and length scales in $\mathbf{B}_h$ may vary across different forecast ranges. Dipankar et al. (2020) noted that the spinup duration in similar downscaled simulations over the same domain is around 6–9 h. Consequently, the ensemble statistics computed from the 3-h forecasts may only be estimating the large-scale forecast errors. In the global implementation, Clayton et al. (2013) noted only a small difference in the ensemble statistics between various lead times and expected only minor impacts on the verification scores. We investigate these issues further and discuss their relevance in section 3.

d. Localization and weightings

A key aspect of the hybrid En3D-Var algorithm is the design of the localization applied on $\mathbf{B}_c$. In many respects, the localization approach is very similar to Clayton et al. (2013) since SINGV-DA also uses the same data assimilation framework of the UKMO, but for the LAM implementation. Here, we focus on the key differences and other important aspects for consideration in the SINGV-DA implementation.

Lorenc (2003) discusses how localization directly in the space of the model variables (hereafter referred to as model space) can result in the generation of subgeostrophic wind increments when a single height observation is assimilated because the Schur product alters the kurtosis of the covariance curve and its gradient. In this light, Clayton et al. (2013) applied balance-preserving localization by first transforming the ensemble perturbations from model space into control variable space (using balance constraints) in the algorithm. Over the western Maritime Continent (deep tropics), we expect
that adhering to geostrophic balance is relatively unimportant. Furthermore, for a convective-scale system, transient dynamical imbalances inevitably occur because of nonlinear convective processes. We have thus opted to perform model space localization in SINGV-DA, which is the approach used in most EnKF algorithms, despite possible resulting dynamical imbalances in the analysis increments. In this manner, multivariate background error relationships are directly prescribed between model variables by the cross covariances [Eq. (5)]. Note that we have also disabled intervariable localization, which removes cross covariances between model variables. Therefore, we preserve the inherent background error relationships captured by the ensemble (see sections 3c and 3d).

We can define the “square-root” $U^a$ using separate horizontal and vertical transforms for the spatial localization:

$$U^a = U_h^a U_v^a,$$

(11)

where $U_h^a$ and $U_v^a$ are the horizontal and vertical transforms that apply the localization, respectively. For the horizontal localization in the LAM implementation, a homogeneous and isotropic Gaussian correlation $C$ is modeled using a spectral representation of

$$C(r) = \exp \left[ -\frac{1}{2} \left( \frac{r}{s} \right)^2 \right],$$

(12)

where $s$ is the horizontal length scale localization parameter, and $r$ is the horizontal distance between two grid points. Note that Eq. (12) (the default Gaussian expression used in the LAM implementation by the UKMO) differs from Clayton et al. (2013) by a factor of 2 for $s$ in the denominator, so the resulting Gaussian correlation function is broader for the same value of $s$. Additionally, a boundary relaxation is applied using half-cosine functions on the ensemble perturbations so that the values are zero at the lateral boundaries. This ensures that $B_c$ satisfies the same boundary conditions that are built into the design of $B_e$.

A side effect of the horizontal localization is the aliasing of the length scales of the analysis increments onto the localization length scale (due to the Schur product), which may degrade the quality of the analysis. To address this, Clayton et al. (2013) introduced an antialiasing “high-pass” horizontal filter to remove the power from the lower wavenumbers (threshold determined by the localization length scale) and spurious gravity wave activity. However, they noted that the justification for its application depends on the length scales present in the ensemble, since it is somewhat an ad hoc modification. As SINGV-EPS is dynamically downscaled and likely contains significant power in the scales larger than the localization length scale, we have chosen to apply this filter in SINGV-DA.

For the vertical localization, an eigendecomposition of a target vertical localization matrix is used. Only a fixed number of leading eigenvectors are retained to reduce the computational costs. In SINGV-DA, the target vertical localization is constructed using a Gaspari–Cohn correlation function [Eq. (4.10) of Gaspari and Cohn 1999] with $\ln(p)$ as a coordinate, where $p$ is the level-mean pressure stored in $B_c$. A $\ln(p)$ separation parameter controls the vertical localization, in a similar manner as $s$ for the horizontal case. This approach follows Buehner (2005) and uses a vertical coordinate such that the same vertical correlation length scales can be used regardless of model level, even with the varying vertical mesh spacing.

Instead of using a uniform weight between $B_e$ and $B_c$ for all model levels, it is possible to use a vertically dependent weighting between $B_e$ and $B_c$ (Buehner et al. 2013; Clayton et al. 2013). Specifically, above a certain height AGL, they introduce a transition zone where the weighting toward $B_c$ is increased...
approximately linearly to full weight. This allows the background error correlation length scales to gradually adjust to the climatological value due to practical model lid constraints in their setup. Gradually weighting toward $B_c$ also allows the horizontal correlation length scales to vary in the upper model levels (e.g., stratosphere), instead of using a single localization length scale in $B_e$, which would likely be shorter than appropriate for the stratosphere (Clayton et al. 2013) since the horizontal correlation length scales tend to be larger in that region (e.g., using observed residuals, Lömberg and Hollingsworth 1986; Bartello and Mitchell 1992, and using forecast difference statistics, Ingleby 2001). Over the Maritime Continent, the tropopause is expected to be around 16 km in height AGL, so we have chosen the transition zone to be 16–21 km (i.e., increasing weighting toward $B_c$, starting from 16 km).

3. Structures in the ensemble-derived background error statistics

a. Analysis of ensemble perturbation structures

It is helpful to examine the spatial variation of the ensemble perturbations since the analysis increment computation involves a linear combination of the ensemble perturbations and is constrained to be within the subspace spanned by them (Lorenc 2003). To illustrate the spatial variation present in the ensemble perturbations, we plot the ensemble member 1 perturbation fields ($X^i_1$) of the horizontal wind vector, pressure, total specific humidity and potential temperature at model level 15 (~1-km height AGL), valid at 0600 UTC 1 Jun 2019 (6-h forecast from 0000 UTC 1 Jun 2019). The vector represents the horizontal wind deviation from the ensemble mean.

![Fig. 2. Ensemble perturbation fields of the (a) horizontal wind, (b) pressure, (c) total specific humidity, and (d) potential temperature for ensemble member 1 at model level 15 (~1-km height AGL), valid at 0600 UTC 1 Jun 2019 (6-h forecast from 0000 UTC 1 Jun 2019). The vector represents the horizontal wind deviation from the ensemble mean.](image-url)
afternoon (Dipankar et al. 2020; Lee et al. 2021) during intermonsoon seasons. Therefore, the forecast errors tend to be larger over the ocean during the diurnal cycle peak (and immediately after the peak, at 0600 UTC).

Next, we compute the mean power spectrum of the ensemble perturbations across all ensemble members and cycles in June 2019 (11 ensemble members and 2 cycles per day, total of 660 samples) for each lead time (Fig. 3). The mean power spectrum for all four variables shows that for shorter lead times (3-h forecasts), the high-resolution structures in the ensemble perturbations (smaller scales) do not contain the same power as for longer lead times (6-, 9-, and 12-h forecasts). This indicates that the ensemble statistics computed from the 3-h forecasts are estimating mainly the large-scale forecast errors. The differences are much smaller in the spectrum of the ensemble perturbations computed using the 6-, 9-, and 12-h forecasts, which suggests that their forecast error variances do not differ substantially, especially after 9 h. This aligns with the notion that the spinup duration of 6–9 h—following Dipankar et al. (2020)—is sufficient for the high-resolution structures in the forecasts to develop. The mean power spectrum of the ensemble perturbations is also closely related to the mean ensemble spread by definition. A higher power is derived from having larger perturbations from the ensemble mean, and this implies a larger ensemble spread. Having similarities in the mean power spectrum for longer lead times indicate that the increase in ensemble spread after 9 h is fairly muted. Clayton et al. (2013) commented that using an ensemble constantly recentered around a deterministic analysis, they had only small differences in the variances and length scales of the ensemble perturbations, even when comparing across different lead times.

For the mean power spectrum for pressure, the ensemble perturbation fields contain mainly large-scale structures (highest power at the largest total horizontal wavelength of 1600 km). However, for total specific humidity and potential temperature, the highest power occurs mainly at around a total horizontal wavelength of 800 km, which indicates that there are smaller scale ensemble perturbation structures in these fields which are more dominant. Consequently, following the schematic (Fig. 1), it is reasonable to expect that the analyses for some SINGV-DA cycles (0300 and 1500 UTC) would be disadvantaged by the lack of representation or underrepresentation of small-scale forecast errors in the ensemble perturbations, unless time shifting of ensemble perturbations (i.e., using ensemble perturbations that are valid prior and after the target analysis time; Lorenc 2017) is considered.

b. Selection of localization length scales from autocovariance structures

To better understand the localization scales suitable for this En3D-Var setup, we compute the raw $P_f$ (valid 0600 UTC 1 June 2019) with respect to a point near the center of the
SINGV-DA domain, focusing on the model level 15 autoco-
variance structures for each of the four variables. One
would expect larger positive covariances near the point
of interest and negligible covariances distant from the point
of interest, assuming the absence of large-scale phenomena
causing long-range spatial correlations. Figure 4 shows that
when using only 11 ensemble perturbations, the autocovar-
iances around the point of interest are generally dominated
by sampling noise.

We explore recomputing the autocovariances using time-
shifted ensemble perturbations, by including the ensemble
perturbations valid 3 h prior and after the target analysis
time. This also serves to artificially boost the size to a total of
33 ensemble perturbations. Figure 5 shows that when using
33 ensemble perturbations, the autocovariances around the point of interest are generally dominated by sampling noise.

We repeat the computation of the autocovariances (33 en-
semble perturbations) with respect to 10 other (land, ocean
and coastal) points in the domain. These had similar qualitative
takeaways on the localization radius and the similarities in the
autocovariance structures between total specific humidity and
potential temperature (not shown). Note that in most cases,
the sampling noise is drastically reduced, but still prevalent.

In the SINGV-DA implementation, the same spatial locali-
ization is applied to all variables, regardless of horizontal position
or model level. The specific humidity field is chosen as a
benchmark for determining the localization length scales,
since it was statistically the noisiest. However, one can argue
that from Fig. 5, the localization length scales suitable for
other variables (e.g., potential temperature) are also com-
parable. Over the western Maritime Continent, the variation
in the total specific humidity background errors are partly gov-
erned by localized hydrometeor-related processes. Destouches
et al. (2021) estimated the optimal horizontal localization length
scales for hydrometeor variables to be around 20–80 km and

![Figure 4. Raw ensemble-derived autocovariances of (a) horizontal wind (with respect to a southwesterly wind),
(b) pressure, (c) total specific humidity, and (d) potential temperature at model level 15 (−1-km height AGL),
computed from the 11 ensemble perturbations valid at 0600 UTC 1 Jun 2019 (6-h forecast from 0000 UTC 1 Jun 2019),
with respect to a point in the center of the domain (black cross). The vector represents the horizontal wind covarian
ces (i.e., positive covariances in both the zonal and meridional components are represented by a vector pointing
northeast).](image-url)
around 120 km for specific humidity. Together with Figs. 4 and 5 (and considering that we use a smaller ensemble), one can hazard an educated guess on the appropriate horizontal localization length scale, of about 50 km, to eliminate the detrimental impacts of sampling noise yet retain meaningful spatial information. This value is also similar to the climatological background error correlation length scales for specific humidity. For vertical localization, we also take reference from Destouches et al. (2021) and climatological background error statistics—the optimal vertical localization length scale, in units of ln(p), for specific humidity using a Gaussian function is around 0.5. A comparable separation parameter for a Gaspari–Cohn correlation function is around 1.5 [i.e., no correlation beyond a ln(p) separation of 1.5]. The selection of localization length scales based on climatological background error statistics is not new; Clayton et al. (2013) also selected the horizontal localization length scales based on error correlation length scales for streamfunction. A more robust and objective approach has previously been developed (Ménétrier 2015a,b), using sample estimated quantities and a linear filter. It is worth considering to reselect the length scales using such advanced system-specific methods once the development of SINGV-EPS becomes more mature. We provide further comments in section 4e.

c. Potential temperature pseudo-observation

To illustrate the effect of the chosen localization parameters, we insert a single pseudo-observation of potential temperature (1 K above the background) near the center of the domain at different model levels and assess the resulting analysis increments. This variable choice also allows us to further investigate the existence of a multivariate relationship between the background errors of total specific humidity and potential temperature highlighted in sections 3a and 3b.

Figure 6 shows the vertical cross section of the potential temperature and total specific humidity responses to the single pseudo-observation of potential temperature (1 K above the background) near the center of the domain at different model levels and assess the resulting analysis increments. This variable choice also allows us to further investigate the existence of a multivariate relationship between the background errors of total specific humidity and potential temperature highlighted in sections 3a and 3b.
Fig. 6. Total specific humidity analysis increment response to a pseudo–single observation of potential temperature 1 K above the background (observation error of 0.5 K) inserted near the center of the domain (green cross) at model levels 15, 29, 40, and 49 (shown in rows from top to bottom, corresponding to 1-, 4-, 8-, and 12-km height AGL, respectively) for (left) 3D-Var, (center) pure En3D-Var without vertical localization, and (right) pure En3D-Var with full spatial localization. See text for details on ensemble perturbations used in pure En3D-Var.
Using 3D-Var, there is a strong drying associated with the positive potential temperature pseudo-observation when it is inserted at the lower model levels (i.e., 15 and 29). When inserted at higher levels, the total specific humidity increments are negligible. This relationship is largely dependent on the calibrated nonlinear transformed humidity control variable (Ingleby et al. 2013) which for SINGV-DA, prescribes strong negative background error cross covariances between potential temperature and total specific humidity when the background is relatively far from saturation in the lower troposphere. This relationship is also captured when using pure EnVar, although the covariance is much smaller and more localized. Moisture is occasionally added at locations adjacent to its removal. When the pseudo-observation is inserted at higher model levels, moisture may be added or removed depending on insertion height AGL, which could be a reflection of the actual day-to-day variability of the background error statistics.

d. Time-averaged cross correlation with potential temperature

We also compute the raw cross correlation between the total specific humidity and potential temperature, using ensemble perturbations from 6-h forecasts sampled from all ensemble members and cycles in June 2019, for different model levels and locations (Fig. 7). It is evident from Fig. 7 that there exists a moderate negative correlation (~0.4) between the two variables at the lower levels, subject to small spatial variations. This correlation weakens higher up in the troposphere. At model level 49 (12 km), the correlation is occasionally weakly positive (~0.2) instead, albeit extremely localized. This was also noted for cross correlations with respect to other points in the domain. Our findings on the negative correlation in the lower troposphere is consistent with Lorenc (2007), who analyzed radiosonde innovation statistics in the UKMO global model. They also found a weak
negative correlation between errors of specific humidity and temperature in the lower troposphere and postulated that this was associated with the presence or absence of clouds, highlighting incorrect vertical motion (e.g., excessive descent leading to warming and drying) as a possible root cause. Over the Maritime Continent, where tropical convection and hydrostatic imbalance is prevalent, incorrect vertical motion (including positional errors in convection) is likely a dominant source of background error. Additionally, moisture convergence leading to vertical motion is usually restricted to the lower troposphere, which explains why the negative correlation is mainly confined below 8 km. The results in sections 3c and 3d indicate that there exist potentially meaningful multivariate background error correlations. While enabling intervariable localization removes sampling noise between variables, it also inadvertently removes these background error correlations. Therefore, it seems reasonable to disable intervariable localization (section 2d) even though the ensemble size is relatively small.

4. Experimental setup and trials

a. Description of trials

The impact of hybrid En3D-Var on SINGV-DA forecasts was evaluated over June 2019, which featured both localized thunderstorms and large-scale convective occurrences throughout the domain. The initial development work trialed hybrid En3D-Var in SINGV-DA using the 11-member SINGV-EPS (Fig. 1) with different weightings between $B_\alpha$ and $B_\beta$ (Table 1). An additional experiment (EXPT-100C-80E) was included to trial using substantial inflation of the background error statistics, augmenting a fully weighted $B_\beta$ with flow-dependent information from $B_\alpha$. Separate tests showed that SINGV-EPS can be underdispersive, especially during the peak of the diurnal cycle. Additionally, $B_\beta$ variances were originally obtained from the lagged NMC method without any tuning. Thus, EXPT-100C-80E helps to preliminarily assess if larger background error variances in general are desirable for SINGV-DA. We also performed two additional experiments incorporating time-shifted ensemble perturbations based on two of the weighting combinations. In all experiments, where required, the localization settings follow the description and justification in sections 2d and 3b.

b. Impact on analysis increments

To illustrate the effect of varying the weighting, Fig. 8 shows the model level 29 (~4 km) horizontal cross section of the potential temperature, total specific humidity, and wind analysis increments for the first cycle of the monthlong trials (0300 UTC 1 June 2019). The same first guess has been used in all experiments.

The analysis increments structures in general appear as a combination of the experiments with 100% $B_\alpha$ or 100% $B_\beta$ (i.e., CTRL or EXPT-0C-100E, respectively), which is logically expected. We note that for experiments with higher weightings of $B_\alpha$, the analysis increments contain smaller scale structures. This was also previously highlighted in Montmerle et al. (2018). Localized values of potential temperature increments, for example, are slightly larger over certain regions (e.g., off the east coast of the Malaysian Peninsula), associated with larger local variances reflecting the ensemble forecast uncertainty over those regions. Most of the analysis increments at this level are due to satellite and aircraft observations, since no radiosonde information is available for this particular cycle.

c. Verification against conventional observations

Next, we assess the background fit to conventional observations by computing the root-mean-square differences between the observation and background (O-B RMS), averaged over all cycles during the monthlong trials. In most of the experiments, there was a reduction in the O-B RMS for radiosonde relative humidity, zonal and meridional wind, and aircraft zonal wind compared to CTRL (Fig. 9). The vertical profiles of differences in O-B RMS compared to CTRL for radiosonde relative humidity, zonal and meridional wind (Fig. 10) also highlight this reduction, particularly below model level 29 (~4 km). For EXPT-0C-100E, the forecast fit to all conventional observations was poorer. This is likely a consequence of applying strict localization in a data-sparse region such as over the western Maritime Continent, so the analysis increments are very localized and observational information is not well spread throughout the domain by $B_\beta$ (Fig. 8; rightmost panel).

In all the experiments, there was also an increase in the O-B RMS for surface temperature and relative humidity compared to CTRL. An increase in the weighting toward $B_\beta$ appears to result in larger O-B RMS values. The vertical profiles of O-B RMS for radiosonde temperature and relative humidity also show that the O-B RMS is larger near the surface compared to CTRL. Note that this result is not seen in the wind variables; we do not assimilate wind information.

---

TABLE 1. Summary of SINGV-DA configurations testing hybrid En3D-Var with different weightings to climatological and ensemble components, and application of time-shifted ensemble perturbations.

<table>
<thead>
<tr>
<th>Experiment name</th>
<th>Weightings</th>
<th>Time-shifted ensemble perturbations</th>
</tr>
</thead>
<tbody>
<tr>
<td>CTRL (3D-Var)</td>
<td>100% $B_\beta$</td>
<td>No (11 ensemble perturbations)</td>
</tr>
<tr>
<td>EXPT-80C-20E</td>
<td>80% $B_\beta$, 20% $B_\alpha$</td>
<td>No (11 ensemble perturbations)</td>
</tr>
<tr>
<td>EXPT-50C-50E</td>
<td>50% $B_\beta$, 50% $B_\alpha$</td>
<td>No (11 ensemble perturbations)</td>
</tr>
<tr>
<td>EXPT-20C-80E</td>
<td>20% $B_\beta$, 80% $B_\alpha$</td>
<td>No (11 ensemble perturbations)</td>
</tr>
<tr>
<td>EXPT-0C-100E (Pure EnVar)</td>
<td>100% $B_\beta$</td>
<td>No (11 ensemble perturbations)</td>
</tr>
<tr>
<td>EXPT-100C-80E</td>
<td>100% $B_\beta$, 80% $B_\alpha$</td>
<td>No (11 ensemble perturbations)</td>
</tr>
<tr>
<td>EXPT-80C-20E-TS</td>
<td>80% $B_\beta$, 20% $B_\alpha^{b^b}$</td>
<td>Yes (33 ensemble perturbations)</td>
</tr>
<tr>
<td>EXPT-50C-50E-TS</td>
<td>50% $B_\beta$, 50% $B_\alpha^{b^b}$</td>
<td>Yes (33 ensemble perturbations)</td>
</tr>
</tbody>
</table>

---

Unauthenticated | Downloaded 06/09/24 07:27 PM UTC
from surface observations because the prevailing wind in the tropics is weak and noisy. One possible reason for the poorer fit to observations near the surface may be related to the inconsistencies in the ECMWF soil moisture data and the Unified Model soil moisture scheme used in SINGV-EPS, which would impact the ensemble perturbations used in hybrid En3D-Var. This technical change will be explored in future studies.

d. Verification against satellite-derived precipitation

To assess the impact of hybrid En3D-Var on SINGV-DA precipitation forecasts, we compute the fractions skill score (FSS; Roberts and Lean 2008) statistics from hourly accumulated precipitation of 60 forecasts (in June 2019) initialized at 0300 and 1500 UTC over the SINGV-DA domain. We have followed the routine verification procedure of the SINGV-DA system, which focuses on 0300 and 1500 UTC because the latest sets of ECMWF LBCs are available only for these two cycles (prior to availability four times a day). We also first focus on the Singapore radar domain since we expect that the impact of hybrid En3D-Var will be more pronounced in regions with more observations (such as in the vicinity of Singapore), given the strict localization. The FSS are computed using a neighborhood size of 50 km, as a function of eight precipitation thresholds (0.125, 0.25, 0.5, 1, 2, 4, 8, and 16 mm). The verification is performed against the Global Precipitation Mission (GPM) data created with the Integrated Multi-satellite Retrievals for GPM, Final Precipitation product (GPM_3IMERGHH v06B, Huffman et al. 2019), which is available at 0.1° × 0.1° spatial resolution, and a 30-min temporal resolution.

Over the Singapore radar domain, there is generally an improvement in the precipitation forecasts compared to CTRL in EXPT-80C-20E, EXPT-50C-50E, and EXPT-100C-80E, up to a lead time of 21 h, particularly for thresholds above 2 mm (Fig. 11). Increasing the weighting toward $B_c$ (including full weight in EXPT-0C-100E) led to larger degradations in the very short-range forecasts, reflecting the immediate effect of the data assimilation. This is likely due the poorer fits to conventional observations, particularly near the surface, as seen in Fig. 9. Incorporating substantial inflation, as done in EXPT-100C-80E, also improved the 9–21-h forecasts, but degraded the forecasts at longer lead times. Overall, EXPT-80C-20E yielded the largest forecast improvements, which is unsurprising, following Hamill and Snyder (2000) who suggested that the optimal combination should be weighted toward $B_c$ if the ensemble size is small. Previous studies (e.g., Lorenc 2003; Ménétrier et al. 2015a,b) have also noted that the suitability of localization length scales depends on ensemble size. We found that in this case, our choice of the strict localization length scales appears to be appropriate given the small SINGV-EPS ensemble size.
Over the full domain, the precipitation forecasts are improved in EXPT-50C-50E and EXPT-20C-80E compared to CTRL for thresholds above 4 mm, but degraded at very short lead times for thresholds below 4 mm, as also seen in Fig. 11. The results for EXPT-80C-20E and EXPT-100C-80E are mixed, with very small degradation or improvements in the forecasts across different thresholds and lead times. The differences between Figs. 11 and 12 suggest that while the forecasts are improved over the Singapore radar domain, the forecasts may be degraded over other parts of the domain, where sampling noise may be more prevalent (especially around Sumatra; not shown) despite the strict localization.

Sections 3b and 3c showed that using time-shifted ensemble perturbations reduces the sampling noise in $B_c$ and highlighted a robust negative background error correlation between total specific humidity and potential temperature. To assess the impact of using time-shifted ensemble perturbations on the precipitation forecasts, we compare the FSS for EXPT-80C-20E-TS and EXPT-50C-50E-TS with EXPT-80C-20E and EXPT-50C-50E, respectively. Figure 13 shows that when time-shifted ensemble perturbations are used, the precipitation forecasts over the whole domain are generally improved. Over the Singapore radar domain, the impact is neutral; EXPT-80C-20E-TS and EXPT-50C-50E-TS yield relatively equal improvements as their experiment counterparts without time-shifted ensemble perturbations, with respect to CTRL (not shown). The improvements when using time-shifted ensemble perturbations agree with the results in section 6.2.2 of Gustafsson et al. (2014). They commented that the use of time-shifted ensemble perturbations allows for timing errors and spatial phase errors to be represented, which perhaps is critical over convection-dominated regions, such as the western Maritime Continent.

e. Other experiments and discussion

We have also conducted further experiments that are not included in the results above. As alluded to in section 2a, the training data for calibrating $B_c$ was generated using a 4.5-km forecast system, whereas SINGV-DA now uses a 1.5-km horizontal grid spacing. One would expect the forecast errors (and their variances) to be smaller in a higher resolution system, hence we tested reducing the standard deviation of streamfunction and velocity potential in $B_c$ by 0.8, together with hybrid En3D-Var. This led to further improvements in the precipitation verification scores and forecast fits to conventional observations, and was thus included in subsequent package upgrade trials.

In section 3b, we noted that there were other objective methods for selecting localization length scales, although the approach by trial-and-error is often adopted by many operational centers. We have hence conducted further experiments using different horizontal localization length scales of up to 200 km with various weightings. The details are omitted in this article given the large number of possible permutations. None of the other experiments outperformed EXPT-80C-20E-TS and in general, given the small SINGV-EPS ensemble size, stricter localization yielded better precipitation verification scores.

A common criticism of using time-shifted ensemble perturbations is that the ensemble perturbations are, strictly speaking, not...
independent samples. One would expect that by omitting cross covariances between ensemble perturbations, the full ensemble-derived covariances would be underestimated, and inflation of the resulting covariances may be required. We have nonetheless opted to test the time-shifting approach for our system, following Gustafsson et al. (2014), Huang and Wang (2018) and Gasperoni et al. (2022). Huang and Wang (2018) had also previously shown that using time-shifted ensemble perturbations improved the Gaussianity of the background ensemble distribution, because time-shifting produces a temporal smoothing effect on the ensemble-derived covariances. They focused on a tropical cyclone case, but for the western Maritime Continent the time-shifted ensemble perturbations may contain diurnal convection signals at different locations, and thus may lead to non-Gaussianity.

5. Conclusions

At the Meteorological Service Singapore (MSS), a hybrid ensemble-variational (En3D-Var) data assimilation system has been developed to explore incorporating information from an ensemble prediction system into a variational data assimilation system over the western Maritime Continent. In this initial implementation, the 11-member ensemble prediction system is dynamically downscaled from global ensemble members every 12 h, with ensemble-derived background error statistics used in the 3-hourly cycling variational data assimilation system.

To understand the ensemble-derived background error statistics, we analyzed the structures and raw covariances of the flow-dependent ensemble perturbations from the ensemble prediction system. There exists small-scale error structures associated with positional differences of tropical convection, but these structures are well represented only after the downscaled ensemble forecast has evolved for at least 6 h due to spinup. This result highlighted a limitation of the initial implementation of the hybrid En3D-Var system, where certain cycles were disadvantaged by the underrepresentation of small scale forecast errors in the ensemble perturbations. Sampling noise was prevalent in the raw autocovariances computed using such a small ensemble, with an estimated horizontal localization scale of around 50 km suitable for this setup. We found that time shifting of the ensemble perturbations, by using those available from adjacent cycles, helped to ameliorate sampling error.

We also discovered a robust and moderate negative correlation between total specific humidity and potential temperature background errors which was confined to the lower troposphere. The negative covariance was also captured by the control variable transform in 3D-Var, but when using the alpha control variable transform with the ensemble-derived covariance, the relationship was weaker and more localized. We postulate that this robust relationship is associated with incorrect vertical motion in the presence of clouds.

Monthlong trials in June 2019 were conducted to assess the impact of hybrid En3D-Var on the analysis increments,
forecast fits to observations and precipitation forecasts. Multiple trials were conducted using different weights assigned to the ensemble-derived and climatological background error covariances, respectively. The analysis increments generally contained smaller scale structures and had larger localized values reflecting the larger forecast uncertainty over certain regions as the weighting toward the ensemble-derived background error covariances increased. The forecast fits to radiosonde relative humidity and wind observations were generally improved with hybrid En3D-Var, but in all experiments, the forecast fits to surface temperature and relative humidity observations were degraded compared to the baseline 3D-Var configuration. Over the Singapore radar domain, there was a general improvement in the precipitation forecasts, particularly for thresholds above 2 mm, compared to the baseline 3D-Var configuration, especially when the weighting toward the climatological background error covariance was larger (e.g., 50% or 80% weight). However, the results were mixed over the full domain, possibly because sampling noise was more prevalent over other parts of

**Fig. 11.** Hinton diagrams of fraction skill scores (FSS) computed over the Singapore radar domain (red rectangle in top-right panel) for all experiments (without time-shifted ensemble perturbations) with respect to CTRL, verified against GPM data. A green (purple) triangle indicates that the forecasts are improved (degraded). A larger triangle indicates a greater improvement or degradation, by up to 0.08 (the same size as the bounding box). Significance is determined using the nonparametric two-sided Wilcoxon signed-rank test at the 90% confidence level, indicated using bold triangles.
the domain (e.g., Sumatra). This issue was mediated by the application of time-shifted ensemble perturbations, which then led to an improvement in the precipitation forecasts instead. Overall, the experiment using a weighting of 80% climatological, 20% ensemble-derived background error covariances, with time-shifted ensemble perturbations, yielded the best verification scores. These results are encouraging, given the simple initial implementation where SINGV-EPS is not centered on the SINGV-DA analysis and is uninformed of the SINGV-DA observation network.

Future work involves consolidating the ensemble prediction system and the deterministic system by centering the ensemble prediction system on the hybrid analysis. This should avoid spinup issues and better represent the analysis and forecast uncertainties since in the consolidated system, the 3-h ensemble forecasts centered on the SINGV-DA analysis are available for all cycles. The ensemble analysis perturbations can also be generated using various ensemble approaches (e.g., bred vectors, ensemble of 3D-Var).

It would also be interesting to explore the error structures in the ensemble perturbations during other seasons, such as the northeast monsoon. Previous investigations have identified other error structures reminiscent of a sea breeze off the coast of Sumatra, but these were present only after full onset of the southwest monsoon (in July; not shown, and in September; Lee and Huang 2022). It would be interesting to see if capturing this flow-dependent information provides the same benefit on precipitation forecasts.

![Fig. 12. As in Fig. 11, but over the full domain (red rectangle in top-right panel).](image)
We are also considering reducing the ensemble horizontal grid spacing from 4.5 to 2.2 km in the future. This brings the current SINGV-EPS horizontal grid spacing closer to that in SINGV-DA. Currently, ensemble forecasts are interpolated from 4.5- to the 2.83-km SINGV-DA variational assimilation grid, which may introduce unintended smoothing effects. However, there is also no guarantee that reducing the ensemble horizontal grid spacing will drastically improve the analysis. Feng and Wang (2021) previously showed that there was a larger positive impact on the analysis when the horizontal grid spacing of the first guess (from the deterministic system) compared to the ensemble forecasts is reduced. Therefore, we may instead choose to reduce the horizontal grid spacing of the first guess in further trials with hybrid En3D-Var.

Acknowledgments. We thank Xiang-Yu Huang for the initial suggestion to develop this system, and Ross Bannister and Javier Amezcua for providing constructive comments to improve the clarity of the manuscript. We would also like to thank the three anonymous reviewers for their suggestions. The authors declare no conflict of interests.

Data availability statement. Due to confidentiality agreements, supporting data can only be made available to researchers subject to a nondisclosure agreement. Queries on details of the data and procedure for requesting access may be directed to the corresponding author.

REFERENCES


Buizza, R., and D. Richardson, 2017: 25 years of ensemble forecasting at ECMWF. ECMWF Newsletter, No. 153, ECMWF, Reading, United Kingdom, 20–31, https://www.ecmwf.int/node/18198.


