Using satellite data assimilation techniques to combine infrasound observations and a full ray-tracing model to constrain stratospheric variables

Javier Amezcua[a, b] , Sven Peter Näsholm[c,d] , Ismael Vera-Rodriguez[e,f]

[a] *Tecnologico de Monterrey Campus Ciudad de Mexico, School of Science and Engineering, Mexico*

[b] *Department of Meteorology, University of Reading, United Kingdom*

[c] *NORSAR, Gunnar Randers vei 15, Kjeller, Norway*

[d] *Department of Informatics, University of Oslo, Norway*

[e] *Silixa, 3102 W Broadway St., Missoula MT, USA*

[f] *formerly at NORSAR, Gunnar Randers vei 15, Kjeller, Norway*

Corresponding author: Javier Amezcua, j.amezcua@reading.ac.uk

1

ABSTRACT: Infrasound waves generated at the Earth's surface can reach high altitudes before returning to the surface to be recorded by microbarometer array stations. These waves carry information about the propagation medium, in particular, temperature and winds in the atmosphere. It is only recently that studies on the assimilation of such data into atmospheric models have been published. Intending to advance this line of research, we here use the Modulated Ensemble Transform Kalman Filter (METKF) –commonly used in satellite data assimilation– to assimilate infrasound-related observations in order to update a column of three vertically varying variables: temperature and horizontal wind speeds. This includes stratospheric and mesospheric heights, which are otherwise poorly observed. The numerical experiments on synthetic data but with realistic reanalysis product atmospheric specifications (following the Observing System Simulation Experiment paradigm) reveal that a large ensemble is capable of reducing errors, especially for the wind speeds in stratospheric heights close to $30 - 60$ km. While using a small ensemble leads to incorrect analysis increments and large estimation errors, the METKF ameliorates this problem and even achieves error reduction from the prior to the posterior mean estimator.

SIGNIFICANCE STATEMENT: The stratosphere and mesosphere have significantly less observational coverage compared to the troposphere, especially for the winds. This lack of information can reduce the accuracy of medium-range weather forecasts. By mimicking a realistic setup, this study paves the way for including novel observations in the estimation of the atmospheric state in these heights using an ensemble data assimilation method. These observations come from a dataset of opportunity containing infrasound-related measurements that are routinely carried out at several stations around the world.

1. Introduction

The skill of numerical weather prediction (NWP) forecasts has increased dramatically over the last 50 years. Bauer et al. (2015) labelled this improvement "a quiet revolution", and it is due to three main factors. First, the dynamical models which describe and predict the evolution of the atmospheric flow have improved considerably (Williamson 2007; Lynch 2008). These have evolved in tandem with increasing computational power, which is needed as the spatio-temporal resolution increases and more processes are represented explicitly. Furthermore, the modeling and forecasting paradigm has shifted from single-trajectory to ensemble systems, which allows us to equip the forecast with a probabilistic measure (Palmer 2019). The second factor is the availability of observations, both in the temporal and spatial domains. The adequate use of remote-sensing observations has proved vital to initialise and constrain NWP forecasts (Bluestein et al. 2022), like the correct usage of the ever-expanding constellation of Earth-observation satellites (Eyre et al. 2020, 2022). Finally, we have the ever-evolving data assimilation (DA) techniques whose purpose is to optimally combine information from these two (imperfect) sources of information (e.g., Asch et al. 2016).

Today, NWP faces novel challenges and new frontiers in forecasting. Research on sub-seasonal to seasonal forecast models has grown over the last decade (Vitart and Robertson 2018). This is a challenging endeavor, since these processes exist somewhere between weather and climate phenomena, and they have components associated both with the initial-value problem aspects of NWP and the boundary-condition aspects of climate prediction. Predictability on these timescales is closely related to the memory of the oceans. Furthermore, the stratosphere and its connection to the troposphere also play a fundamental role in this regime (Baldwin and Dunkerton 2001;

3

Baldwin et al. 2003). There is great potential in extending the assimilation of stratospheric data with appropriate accuracy, in particular for variables like wind velocity, which are difficult to observe. For instance, the European Centre for Medium-Range Weather Forecasts (ECMWF) has detected systematic model errors in the stratosphere over short-range forecasts and reanalyses (Shepherd et al. 2018; Simmons et al. 2020), and attempted to compensate for such errors using machine learning techniques (Bonavita and Laloyaux 2020).

Recent years have seen a gradual increase in the availability of observations for the middle atmosphere (stratosphere and mesosphere). In 2018, the European Space Agency (ESA) launched the Aeolus satellite, which carried the first Doppler Wind Lidar (DWL) in space. In January 2020, ECMWF started assimilating atmospheric winds retrieved by this instrument (Rennie et al. 2021), including both Rayleigh-clear and Mie-cloudy winds. The instrument was designed to provide information on the wind from the surface up to the lower stratosphere (~ 30 km). Figure 1 of Rennie et al. (2021) shows an example of a retrieval up to $20 - 28$ km. These observations promise to provide improvements in the forecast. For instance, simulation experiments with the ECMWF system have shown error reductions of up to 2% for horizontal winds and temperature on days 2 to 3 in the tropical upper troposphere, and reductions of up to 1% in the polar regions out to several days. Unfortunately, Aeolus deorbited and re-entered Earth's atmosphere in 2023.

It is promising that the observation density in the middle atmosphere is slowly increasing, but there are existing observation networks which are currently not exploited in operational DA. We refer, in particular, to data collected from infrasound station networks around the world (Assink et al. 2019; Marty 2019). These datasets contain information, albeit indirect, of stratospheric variables. It is timely to further explore the use of these data sources to improve the stratospheric representation in atmospheric models and NWP.

This paper exploits an infrasound dataset of opportunity using an ensemble DA method. These data encompass signatures of infrasound from controlled explosions where the propagating waves have entered the stratosphere and reflected back towards the ground. They are then recorded by specialized sensor arrays (see Gibbons et al. 2015; Blixt et al. 2019, as well as Section 2b below). We base our approach on one used to assimilate satellite observations, since these share many similarities with infrasound measurements, since they are both integrated measurements. Compared to previous works, this paper has two main contributions. The first is the integration

of a full ray-tracing model into the DA process instead of geometric models. This complicates the localization step in the ensemble DA process, since a straightforward formulation requires an expensive computation of Jacobians of the ray-tracing model. In the case of radiative transfer models needed for satellite DA (Campbell et al. 2010), this issue has been largely overcome by a handful of methods, one being the Modulated Ensemble Transform Kalman Filter (METKF) introduced by Bishop and Hodyss (2009), and refined by Bishop et al. (2017). METKF is a specialized implementation of the Ensemble Transform Kalman Filter (ETKF) (Bishop et al. 2001) that increases the size of an ensemble of forecasts into a modulated ensemble which, by design, yields a desired localized covariance matrix. The second contribution of this work is the application of the METKF for the assimilation of infrasound-related observations.

This paper follows the paradigm of Observing-Systems Simulation Experiments (OSSE), described in Arnold and Dey (1986). OSSEs are commonly used in the atmospheric science community to assess the potential of assimilating new types of observations. OSSEs are constructed by generating a known synthetic truth (by running an atmospheric model, for instance), simulating imperfect observations for which the observation operator and error are known and characterized, and attempting to reconstruct the original truth from these observations using DA techniques. In the current work, we are interested in testing: (i) the ability of METKF to estimate a vertical atmospheric profile from infrasound-related experiments with a relatively large ensemble (the idealized setting); and (ii) the capacity of the METKF to overcome the sampling problems that are due to limited-size ensembles (the realistic setting).

The rest of this paper is organized as follows. In Section 2 we discuss sources of infrasound data, in particular the Hukkakero-Arces geographical setting used in our OSSE's. Section 3 includes a detailed description of the problem in terms of DA elements, as well as the justification for using the METKF to localize vertically integrated quantities. Section 4 describes the setup for our OSSE's, and Section 5 presents and analyzes the results. We finish with conclusions and discussion in Section 6.

5

2. Infrasound-related measurements and their potential in DA

a. Infrasound propagation as a means to probe the atmosphere

There are different sources at the Earth's surface (and aloft) that generate infrasound waves, that is, acoustic inaudible waves in the $0.1 - 10$ Hz range. These sources of infrasound can be natural (e.g., volcanoes, earthquakes, ocean-generated microbaroms), or the result of human activities (e.g., controlled detonations of munition or quarry explosions). At such low frequencies, these waves can travel for long distances, suffering little dissipation, and having a wide range of propagation characteristics and various sensitivities to atmospheric conditions (Le Pichon et al. 2010, 2019). Depending on the atmospheric structure, some of these waves travel up to middle atmospheric or lower thermospheric altitudes from where they are reflected or refracted back to the surface, to be detected by sensor arrays. While traveling through an atmospheric slab, they are affected by the atmospheric conditions, in particular, winds and temperature. Therefore, the infrasound-related measurements estimated at the instrument site contain useful integrated information about the conditions along the propagation path. For example, an objective of the EU-funded ARISE (Atmospheric dynamics Research InfraStructure in Europe) projects was to establish and ensure the operation of this type of infrastructure for long-term observation and modeling of the middle atmosphere (Blanc et al. 2018, 2019).

In the past, infrasound measurements have been used to evaluate and validate, and invert for atmospheric model profiles. For instance, Assink et al. (2014) used volcanic infrasound to evaluate winds and temperatures of the ECMWF forecast system. Smets et al. (2016) used these measurements to evaluate the forecast skill on predicting sudden stratospheric warming events. In these studies, values coming from an atmospheric profile (either forecast or analysis) are put through a ray-tracing model to generate the *predicted* infrasound-related observations. These results are then compared to the actual observations. This is the so-called *forward* modeling problem. A more challenging problem is to take actual observations and try to infer the atmospheric values to which they correspond, i.e., the so-called *inverse* problem. In general, this is an underdetermined problem, i.e., it does not have a unique solution. Several recent works have explored the inversion of infrasound-related observations to derive atmospheric profiles or to provide corrections to model profiles (e.g., Le Pichon et al. 2005; Drob et al. 2008; Lalande et al. 2012; Johnson et al. 2012;

6

Arrowsmith et al. 2013; Assink et al. 2013; Park et al. 2016). Recently, Amezcua et al. (2020) and Amezcua and Barton (2021) conducted DA experiments using the infrasound dataset described next.

b. Using infrasound observations from controlled detonation of ammunition in data assimilation

We start by noting that, from an atmospheric probing perspective, we are describing a dataset of opportunity. There are several additional repeated explosion datasets available in this region, stemming, e.g., from open-pit mining blasts (Gibbons et al. 2015).

In August and September since 1988, excess explosives have been disposed of by controlled detonations in Hukkakero, Finland. Infrasound waves generated by these detonations are excited in all directions, and their path through the atmosphere is mainly a function of the atmospheric wind and temperature fields. Some of these waves are detected about 10 minutes later by the array station ARCES (Norway), which is ~ 180 km almost due North from the explosion site. It has been concluded that these waves have been partially reflected or back-scattered by smaller-scale atmospheric structures at about $30 - 60$ km altitude (Blixt et al. 2019; Vorobeva et al. 2023). The exact altitude of the strongest reflection is unknown and varies between events, but it can be estimated by indirect methods like matching observed travel times with those of simulated infrasonic wave propagation paths.

This kind of transient explosion infrasound dataset is less complex than many other infrasound datasets since the waves can be approximated to radiate from a single point source in both space and time, with a known location and origin time. These data are sensitive to wind and temperature in the stratosphere and lower mesosphere. A variety of infrasound wave propagation and atmospheric probing studies have exploited infrasound data recorded from these explosions (Gibbons et al. 2015; Green and Nippress 2019; Gibbons et al. 2019; Blixt et al. 2019; Cugnet et al. 2019; Vera Rodriguez et al. 2020; Vorobeva et al. 2023). Blixt et al. (2019) performed near-inversion experiments by associating observed quantities with wind averages over different altitude "channels" (layers) computed from the ERA-Interim product (Dee et al. 2011). They simplified the acoustic transmission problem using a straightforward geometric model. Vera Rodriguez et al. (2020) performed an inversion of these observations applying a heuristic optimization algorithm, using

7

smoothness criteria to regularise the problem, and using the ERA-5 ensemble product (Hersbach et al. 2019) to initialise the process and constrain its output.

Two recently published scientific papers exploited DA for the observations from this data set (Amezcua et al. 2020; Amezcua and Barton 2021). These works applied different flavors of the Ensemble Kalman filter (Evensen 2006, EnKF), a technique widely utilized in atmospheric science, and which can be considered a sample-based implementation of the Kalman filter (Kalman 1960). Both cases used the ERA5 10-member ensemble as background (Hersbach et al. 2019), and the simple geometric model developed by Blixt et al. (2019), using the reflection heights inferred in that work. EnKF relies on sample covariance matrices to spread information from observations to model variables. For the number of variables in the atmospheric profiles at the heart of this problem, the quality of a covariance matrix computed from a 10-member ensemble is low, often plagued with spurious sampling errors. Localization is a way to ameliorate these problems (Hamill et al. 2001; Anderson 2012). This is not trivial when using observations that contain integrated quantities (Campbell et al. 2010). Amezcua et al. (2020) evaded this problem by taking averages and reducing the number of vertical levels to update from 137 to only 6. Amezcua and Barton (2021) used 131 levels and applied a type of localization (Shlyaeva et al. 2019) that required the linearization of the observation operator, which was relatively straightforward for the geometric model (Blixt et al. 2019). However, this is complicated and computationally expensive with a full ray-tracing model, which is needed for a more realistic treatment of infrasound observations.

In the last decade, some methods were developed to implement localization in vertically-integrated satellite observations. For instance, Lei et al. (2016) implemented a global group filter to derive adaptive localization. This filter was similar to the hierarchical filter proposed in Anderson (2007) and later used in Lei and Anderson (2014) to construct vertical localization functions. Farchi and Bocquet (2019) recommend the creation of augmented ensembles by performing e.g. a randomized singular value decomposition. Bishop et al. (2017) developed the Modulated Ensemble Transform Kalman Filter (METKF), a variation of the ETKF (Bishop et al. 2001; Wang et al. 2004) which avoids linearization by computing an extended (modulated) ensemble which has, by construction, a desired localization structure. In the rest of the current paper, we apply this method to assimilate infrasound-related data into realistic atmospheric vertical profiles.

8

3. Methodology

This section introduces basic aspects of the DA method used and presents the nomenclature used in the rest of the work. In this work we focus on a single assimilation step of the DA process, i.e. there is no cycling.

a. Problem setup

Let us consider an atmospheric range-independent volume that is discretized into a profile of N_z vertical levels centered at $\{z_1, z_2, \cdots, z_{N_z}\}$. Each level has N_{var} altitude-dependent physical variables (we do not consider any horizontal variation). In our case, these are the cross-wind $u(z)$, tail-wind $v(z)$ and temperature $T(z)$, rendering $N_{\text{var}} = 3$. At a given time t, the state vector $\mathbf{x} \in \mathbb{R}^{N_x}$ is the concatenation of the $N_x = N_z N_{\text{var}}$ elements. Namely,

$$\mathbf{x} = \left[\mathbf{T}, \mathbf{u}, \mathbf{v} \right]^{\top}, \tag{1}$$

where

$$\begin{aligned}
\mathbf{T} &= \left[T_1, \cdots, T_{N_z} \right]^{\top}, \\
\mathbf{u} &= \left[u_1, \cdots, u_{N_z} \right]^{\top}, \\
\mathbf{v} &= \left[v_1, \cdots, v_{N_z} \right]^{\top}.
\end{aligned} \tag{2}$$

We assume that $\mathbf{x}$ follows a multivariate Gaussian (MVG) distribution, i.e. $\mathbf{x}^b \sim N\left(\boldsymbol{\mu}^b, \mathbf{B}\right)$, where $\boldsymbol{\mu}^b \in \mathbb{R}^{N_x}$ and $\mathbf{B} \in \mathbb{R}^{N_x \times N_x}$ are the background mean and covariance matrix, respectively. If there are unknown system parameters, these can be appended to the state vector and estimated by the DA, using a process called state augmentation (e.g. Asch et al. 2016). In this work, however, we only perform state estimation.

The observations $\mathbf{y} \in \mathbb{R}^{N_y}$ are functions of $\mathbf{x}$, which can be measured. Mathematically,

$$\mathbf{y} = h\left(\mathbf{x}^{\text{true}}\right) + \boldsymbol{\eta}. \tag{3}$$

where $\mathbf{x}^{\text{true}} \in \mathbb{R}^{N_x}$ is the *true* value of the state variable, which is indeed the value we are trying to estimate. The term $\boldsymbol{\eta} \in \mathbb{R}^{N_y}$ corresponds to additive observational error and is assumed to follow an MVG distribution with zero mean (unbiased error). Namely, $\boldsymbol{\eta} \sim N\left(\mathbf{0}, \mathbf{R}\right)$. The observational

9

error covariance $\mathbf{R} \in \mathbb{R}^{N_y \times N_y}$ depends on the measuring system and is often assumed to be known. Finally, the function $h : \mathbb{R}^{N_x} \to \mathbb{R}^{N_y}$ is the so-called observation operator. In remote sensing of atmospheric temperatures using satellite observations of radiance (Campbell et al. 2010), h is a radiation transfer model. The infrasound measurement problem also involves a wave (mechanical instead of electromagnetic) traveling through an atmospheric slab and being modified by the atmospheric conditions it encounters. In both cases, h integrates (or sums) information from state variables, reducing fields into scalars, and yielding $N_y \ll N_x$. Next, we describe how h was previously modeled, and how the modeling has been improved in this work.

Amezcua et al. (2020) and Amezcua and Barton (2021) used a simplified observation operator, derived from the geometrical model of Blixt et al. (2019). Although useful, this is a rough approximation of the real process. It relates the cross-wind $\mathbf{u}$ to the shift in the backazimuth angle δ_{ba} detected at reception, but excludes the tail-wind $\mathbf{v}$ and temperature field $\mathbf{T}$ components included in Eq (1). It considers the transmission path to be a perfect triangle, which is a rough approximation (see, e.g., Ostashev and Norris (2022) to appreciate the real complexity of the ray paths). It also considers a unique reflection height $Z_{\max}$, corresponding to a single reading of the infrasound array processing output time series. In reality, there are typically multiple arrivals corresponding to multiple reflection heights and infrasound processing output readings (see Blixt et al. 2019, Figs. 4 and 5 for illustrations).

Following Vera Rodriguez et al. (2020), in this work we use a full ray-tracing model, InfraGA, developed by Dr. Philip Bloom at the Los Alamos US National Laboratory (Blom et al. 2015; Blom and Waxler 2017, 2021). InfraGA simulates the infrasonic wave propagation from source to receiver, and its numerical modeling links the three atmospheric fields in Eq (2), to three observables estimated using the infrasound data recorded at the array station, which are related to the direction-of-arrival of the infrasound wavefront impinging the station. They are δ_{ba}, t_{travel} (travel time), and v_{trace} (trace velocity). In DA terms, the observation vector is:

$$\mathbf{y} = \left[\delta_{ba}, t_{\text{travel}}, v_{\text{trace}} \right]^{\top}. \tag{4}$$

When applying this InfraGA, we still consider a single and known $Z_{\max}$. As in Blixt et al. (2019), we introduce the reflection in the horizontal middle point of the path. We acknowledge that there is an inherent time-evolution component in h since the infrasound takes a finite time to travel

between the source and the receiver, so the wave encounters a time-evolving atmosphere. In the satellite case, the electromagnetic wave travels at the speed of light, so the atmospheric variability is negligible, but the sound speed is much lower. Nonetheless, and in accordance with previous works, we consider the travel time of the infrasound wave (about 10 minutes in our setting) to be very small compared to the variability time scales of the atmospheric flow.

b. The ensemble transform Kalman filter

Having the background information and observations just described, the goal of DA is to obtain a posterior distribution for the state variable. If all errors involved in the problem are MVG, and $h = \mathbf{H} \in \mathbb{R}^{N_y \times N_x}$ (linear), the posterior distribution (also known as analysis) is also MVG, with mean $\boldsymbol{\mu}^a \in \mathbb{R}^{N_x}$ and covariance $\mathbf{A} \in \mathbb{R}^{N_x \times N_x}$. These values can be obtained using the analysis equations of the Kalman filter (Kalman 1960). The extended Kalman filter (EKF) is an adaptation of the Kalman filter for nonlinear setups, yielding a suboptimal yet useful solution to the problem as long as h is not too far from linear (e.g., Jazwinski 1970). Ensemble Kalman filters (EnKFs) (e.g., Nerger et al. 2012; Houtekamer and Zhang 2016) are sample-based implementations of the EKF, which avoid the computation and storage of the Jacobian matrix of h (required by the EKF).

EnKFs rely on an ensemble to estimate means and covariances. Fortunately, modern ensemble prediction systems (e.g., Palmer 2019) generate N_e forecasts from different initial conditions (selected to represent the uncertainty in the system) and evolved using a forecast model. At a time t, we denote the background ensemble as $\mathbf{X}^b \in \mathbb{R}^{N_x \times N_e}$, where each column is a valid forecast (we do not need a time index since we do not perform cycling). The ensemble mean $\bar{\mathbf{x}}^b \in \mathbb{R}^{N_x}$ and a matrix of perturbations $\hat{\mathbf{X}}^b \in \mathbb{R}^{N_x \times N_e}$ can be written as:

$$\mathbf{X}^b = \left[\mathbf{x}_1^b, \mathbf{x}_2^b, \cdots, \mathbf{x}_{N_e}^b \right]$$
$$\bar{\mathbf{x}}^b = \frac{1}{N_e} \mathbf{X}^b \mathbf{1} \tag{5}$$
$$\hat{\mathbf{X}}^b = \mathbf{X}^b - \bar{\mathbf{x}}^b \mathbf{1}^\top,$$

where $\mathbf{1} \in \mathbb{R}^{N_e}$ is a vector of ones. Note that one can obtain the sample estimator of $\mathbf{B}$, denoted $\mathbf{P}^b$, as:

$$\mathbf{P}^b = \frac{\hat{\mathbf{X}}^b \hat{\mathbf{X}}^{b\top}}{N_e - 1} \tag{6}$$

11

One can define an ensemble of predicted observations $\mathbf{Y}^b \in \mathbb{R}^{N_y \times N_e}$ by applying h to each ensemble member, and compute the corresponding mean $\bar{\mathbf{y}}^b \in \mathbb{R}^{N_y}$ and matrix of perturbations $\hat{\mathbf{Y}}^b \in \mathbb{R}^{N_y \times N_e}$ (e.g., Evensen 1994). Namely,

$$
\begin{aligned}
\mathbf{Y}^b &= \left[h\left(\mathbf{x}_1^b\right), \cdots, h\left(\mathbf{x}_{N_e}^b\right) \right], \\
\bar{\mathbf{y}}^b &= \frac{1}{N_e} \mathbf{Y}^b \mathbf{1} \\
\hat{\mathbf{Y}}^b &= \mathbf{Y}^b - \bar{\mathbf{y}}^b \mathbf{1}^\top.
\end{aligned}
\tag{7}
$$

One can compute the KF equations with the sample elements to transform $\bar{\mathbf{x}}^b$ into $\bar{\mathbf{x}}^a$ and $\mathbf{P}^b$ into $\mathbf{P}^a$ respectively. However, this does not provide a way to transform individual ensemble members. There is no unique way to do this, leading to different EnKF formulations. The ETKF is a deterministic square root filter that uses a right matrix multiplication to transform the perturbations $\hat{\mathbf{X}}^b$ into $\hat{\mathbf{X}}^a$. Namely,

$$
\hat{\mathbf{X}}^a = \hat{\mathbf{X}}^b \mathbf{W}^a,
\tag{8}
$$

where $\mathbf{W}^a \in \mathbb{R}^{N_e \times N_e}$ is referred to as the transform matrix, chosen such that

$$
\mathbf{P}^a = (N_e - 1)^{-1} \hat{\mathbf{X}}^a \hat{\mathbf{X}}^{a\top}
\tag{9}
$$

fulfils the KF analysis equation for covariance. Using the Potter method and the Woodbury matrix inversion lemma, it can be shown (Tippett et al. 2003) that this matrix is:

$$
\mathbf{W}^a = \left(\mathbf{I} + \frac{\hat{\mathbf{Y}}^{b\top} \mathbf{R}^{-1} \mathbf{Y}^b}{N_e - 1} \right)^{-\frac{1}{2}}.
\tag{10}
$$

A matrix $\mathbf{C}^{\frac{1}{2}} \in \mathbb{R}^N$ is a square root of $\mathbf{D} \in \mathbb{R}^N$ if $\mathbf{D} = \mathbf{C}^{\frac{1}{2}} \mathbf{C}^{\frac{1}{2}\top}$. There is an infinite number of square root matrices, but Wang et al. (2004) propose using the symmetric matrix square root:

$$
\mathbf{W}^a = \mathbf{G} \left(\mathbf{I} + \mathbf{\Lambda} \right)^{-\frac{1}{2}} \mathbf{G}^\top
\tag{11}
$$

using the eigendecomposition

$$\mathbf{G}\boldsymbol{\Lambda}\mathbf{G}^{\top} = \frac{\hat{\mathbf{Y}}^{b\top}\mathbf{R}^{-1}\hat{\mathbf{Y}}^{b}}{N_e - 1}, \tag{12}$$

where $\mathbf{G} \in \mathbb{R}^{N_e \times N_e}$ is the matrix of eigenvectors and $\boldsymbol{\Lambda} \in \mathbb{R}^{N_e \times N_e}$ is the (diagonal) matrix of eigenvalues of the $N_e \times N_e$ positive definite and symmetric matrix on the right side of Eq (12). Hunt et al. (2007) note that the mean $\bar{\mathbf{x}}^a$ can be found as the sum of $\bar{\mathbf{x}}^b$ and a linear combination of the columns of $\hat{\mathbf{X}}^b$, i.e.:

$$\bar{\mathbf{x}}^a = \bar{\mathbf{x}}^b + \hat{\mathbf{X}}^b \mathbf{w}^a. \tag{13}$$

The vector $\mathbf{w}^a \in \mathbb{R}^{N_e}$ contains the coefficients for the linear combination. Explicitly:

$$\mathbf{w}^a = \frac{1}{N_e - 1}\mathbf{W}^a\mathbf{W}^{a\top}\hat{\mathbf{Y}}^{b\top}\mathbf{R}^{-1}\left(\mathbf{y} - \bar{\mathbf{y}}^b\right). \tag{14}$$

Finally, $\mathbf{X}^a$ can easily be constructed by adding $\bar{\mathbf{x}}^a$ to each column of $\hat{\mathbf{X}}^a$. Namely,

$$\mathbf{X}^a = \left[\hat{\mathbf{x}}_1^a + \bar{\mathbf{x}}^a, \cdots, \hat{\mathbf{x}}_{N_e}^a + \bar{\mathbf{x}}^a\right]. \tag{15}$$

c. Localization for integrated observation quantities: the modulated ETKF

The sample estimator quality is crucial for EnKF performance. When $N_e \ll N_x$, sampling errors lead to spurious off-diagonal elements in $\mathbf{P}^b$. Recognizing this, both Anderson (2001) and Hamill et al. (2001) proposed to artificially modify $\mathbf{P}^b$ to reduce sampling errors; this modification is called localization.

Localization can be performed either in state space or in observation space, see, e.g., Farchi and Bocquet (2019) for an ample discussion on the topic. The localized covariance $\mathbf{P}_{\text{loc}}^b \in \mathbb{R}^{N_x \times N_x}$ in state space is obtained from the Schur (element-wise) product of $\mathbf{P}^b$ and a localization matrix $\mathbf{L} \in \mathbb{R}^{N_x \times N_x}$. Explicitly, we have

$$\mathbf{P}_{\text{loc}}^b = \mathbf{P}^b \circ \mathbf{L}. \tag{16}$$

A typical choice for this decay is the so-called Gaspari–Cohn function, which is a compact-support approximation to a Gaussian (Gaspari and Cohn 1999). In the case of multiple variables per level, one has to modify the different blocks of the matrix separately. When using model-space localisation, the ETKF formulae discussed previously cannot be applied. A matrix square-root

13

formulation is required instead. This implies that, if the observation operator is nonlinear, one needs to compute its Jacobian as indicated in Shlyaeva et al. (2019). In fact, Amezcua and Barton (2021) applied localization in this way to infrasound observations when using a simple geometric propagation model, since in that case the calculation of the Jacobian was trivial.

Observation-space localization follows a different approach, see, e.g., Hunt et al. (2007) for a detailed description. This paradigm focuses on the variables of a given gridpoint of the model and looks at the observations that affect that gridpoint within a defined radius of interest. The local observation precision (inverse of the covariance) matrix is modified to account for the distance of observations from the gridpoint. This is

$$\mathbf{R}_{\text{i,loc}}^{-1} = \mathbf{R}_{i}^{-1} \circ \mathbf{L}, \tag{17}$$

where $\mathbf{R}_{\text{i}}^{-1}$ is a submatrix of $\mathbf{R}$ specific to the ith gridpoint, and $\mathbf{L}$ is a localization matrix with the corresponding dimensions. The full analysis field is a collection of patches of local analyses, minimizing jumps amongst the patches (Ott et al. 2004). This formulation is completely compatible with the ETKF formulae discussed earlier (Hunt et al. 2007).

Applying localization to integrated quantities with complex observation operators is a challenge. Campbell et al. (2010) analysed the case involving satellite observations. The difficulty arises because the model space is a column with temperatures at each level, whereas the observation space is a set of channels measuring radiance at different wavelengths. The observational data are hence integrated quantities. Therefore, defining a 'distance' between two channels, or between a channel and a vertical level, is not conceivable as required for observation-space localisation. Since in our problem we have spatially integrated infrasound wave propagation data, similar difficulties arise. Our model space is an atmospheric slab with temperature and wind variables at vertical levels, and our observation space has three integrated quantities related to the infrasound propagation through the atmospheric slab. A naïve implementation of model-space localization in our case is not feasible since the Jacobian of the ray-tracing model is not available.

Bishop and Hodyss (2009) and Bishop et al. (2017) developed a clever method to apply localization for integrated observations in the ETKF without requiring a Jacobian. It takes $\mathbf{X}^{b} \in \mathbb{R}^{N_e}$ and

14

converts it into a *modulated* ensemble $\mathbf{U}^b \in \mathbb{R}^{N_m}$, such that both have the same mean:

$$\bar{\mathbf{u}}^b = \bar{\mathbf{x}}^b,\tag{18}$$

but where $\mathbf{U}^b$ produces a localized covariance. This requires a factorization of Eq (16) in the following form:

$$\left(\frac{\hat{\mathbf{X}}^b\hat{\mathbf{X}}^{b\top}}{N_e - 1}\right) \circ \mathbf{L} = \frac{\hat{\mathbf{U}}^b\hat{\mathbf{U}}^{b\top}}{N_e - 1},\tag{19}$$

where $\hat{\mathbf{U}}^b \in \mathbb{R}^{N_x \times N_m}$ is an ensemble of *modulated* perturbations, (hence the name Modulated ETKF). Obtaining $\hat{\mathbf{U}}^b$ from Eq (19) is possible using the Khatri–Rao product (Khatri and Rao 1968). We follow Bishop et al. (2017) for the next steps, albeit with some variations in notation. $\mathbf{L}$ is a positive definite symmetric matrix, allowing for the eigendecomposition:

$$\mathbf{L} = \mathbf{S}\boldsymbol{\Gamma}\mathbf{S}^\top,\tag{20}$$

where $\mathbf{S} \in \mathbb{R}^{N_x \times N_x}$ is the matrix of eigenvectors, and $\boldsymbol{\Gamma} \in \mathbb{R}^{N_x \times N_x}$ is a diagonal matrix of eigenvalues. The non-symmetric matrix square root of $\mathbf{L}$ is:

$$\mathbf{L}^{\frac{1}{2}} = \mathbf{S}\boldsymbol{\Gamma}^{\frac{1}{2}},\tag{21}$$

where $\boldsymbol{\Gamma}^{\frac{1}{2}}$ is trivial to compute since it is diagonal. $\mathbf{L}^{\frac{1}{2}}$ has the same shape as $\mathbf{L}$. In its construction, it is not necessary to keep all the eigenvectors of the localization matrix. Often, the leading N_λ eigenvalues in $\boldsymbol{\Gamma}$ add to 90% (or any threshold) of the total sum of eigenvalues, and many of the smallest eigenvalues are close to zero. Bishop et al. (2017) recommend keeping the leading N_λ eigenvalues to construct a rank-N_λ approximate matrix square root $\tilde{\mathbf{L}}^{\frac{1}{2}} \in \mathbb{R}^{N_x \times N_x}$ as:

$$\tilde{\mathbf{L}}^{\frac{1}{2}} = \mathbf{S}_{[:,1:N_\lambda]}\boldsymbol{\Gamma}^{\frac{1}{2}}_{[1:N_\lambda,1:N_\lambda]}.\tag{22}$$

In Eq (22), the subindices indicate the number of [rows, columns] kept for each matrix. Finally $\hat{\mathbf{U}} \in \mathbb{R}^{N_x \times N_m}$ is computed as the Khatri–Rao product (Khatri and Rao 1968):

$$\hat{\mathbf{U}}^b = \tilde{\mathbf{L}}^{\frac{1}{2}} \diamond \hat{\mathbf{X}}^b.\tag{23}$$

15

The Khatri–Rao product (denoted $\diamond$) is a column-wise Kronecker product. Explicitly, this is:

$$\hat{\mathbf{U}}^b = \left[\left(\ell_1^{\frac{1}{2}} \circ \hat{\mathbf{x}}_1^b, \cdots, \ell_1^{\frac{1}{2}} \circ \hat{\mathbf{x}}_{N_e}^b \right), \left(\ell_2^{\frac{1}{2}} \circ \hat{\mathbf{x}}_1^b, \cdots, \ell_2^{\frac{1}{2}} \circ \hat{\mathbf{x}}_{N_e}^b \right), \cdots, \left(\ell_{N_\lambda}^{\frac{1}{2}} \circ \hat{\mathbf{x}}_1^b, \cdots, \ell_{N_\lambda}^{\frac{1}{2}} \circ \hat{\mathbf{x}}_{N_e}^b \right) \right] \quad (24)$$

We added parentheses to help identify the elements in $\hat{\mathbf{U}}^b$. Each parenthesis contains the original N_e-member ensemble *modulated* by a corresponding eigenvector $\ell_i^{\frac{1}{2}}$ of $\tilde{\mathbf{L}}^{\frac{1}{2}}$. This illustrates that $N_m = N_\lambda N_e$ depends on the number of eigenvalues kept.

The ETKF is applied on $\mathbf{U}^b$, yielding a modulated analysis ensemble $\mathbf{U}^a \in \mathbb{R}^{N_m}$. If one needs to recover an ensemble of the original size (e.g., to continue the prediction process), there are ways to select an ensemble $\mathbf{X}^a$ from $\mathbf{U}^a$ (Bishop et al. 2017), but it is not necessary for us.

4. Applying METKF to synthetic infrasound data: Experimental setup

We now implement the METKF in a synthetic yet realistic case involving infrasound-related observations. We simulate the geographic and atmospheric settings of the Hukkakero explosion site, with infrasound waves propagating to the ARCES array (Blixt et al. 2019). The source and reception sites are separated by an approximate North-South distance of $L \sim 180$ km along the great circle, with infrasound waves reflected roughly between 30 km and 50 km high.

We start the section by describing the construction of a realistic background ensemble $\mathbf{X}^b$. Then, we look into the design of a localization matrix, and the application of h. As is conventional for OSSE experiments, we produce a reference truth $\mathbf{x}^{\text{true}}$ to compare against, and from which synthetic observations are generated. We then apply the ETKF using a large ensemble to get a solution without sampling error. Then we reduce the ensemble size, which introduces sampling error, and finally we ameliorate this applying localization. The performance of the modulated ensemble is compared with a raw ensemble of the same size.

To assess the DA results, we perform two complementary processes. The first is a single-observation experiment. In this case, we take a single truth and generate a 'perfect' observation, i.e., we choose the realization of the observation error to be equal to $E[\boldsymbol{\eta}] = \mathbf{0}$. We then perform DA with this observation and assess the quality of the result. The second is a statistical evaluation of the DA performance. In this case, we consider N_{true} truths and produce their respective observations

16

Background correlation matrix
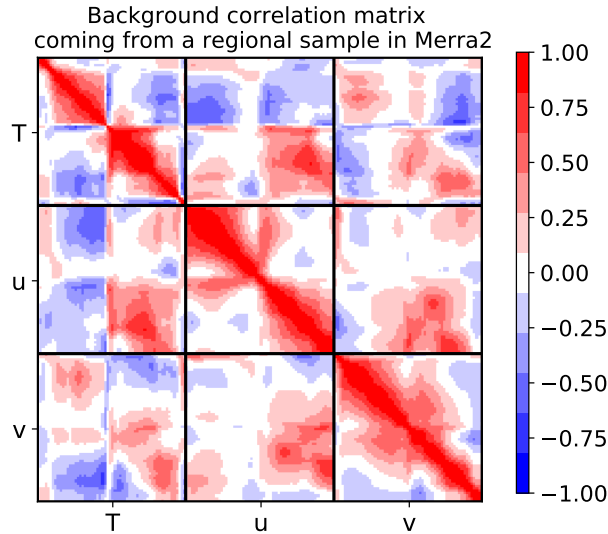coming from a regional sample in Merra2

FIG. 1. Background correlation matrix obtained from a regional sample of MERRA2 data, and used in our experiments. The diagonal blocks correspond to correlations between vertical levels of the same physical variable. The off-diagonal blocks correspond to correlations between different vertical levels of different physical variables.

using different realizations of $\boldsymbol{\eta}$. We perform the DA process, evaluate the performance in each and every case and collect statistics.

a. Background-state generation

Ideally, we would have a large ensemble forecast for a given lead-time in the region of interest. In its absence, we artificially generate a large ensemble that represents realistic atmospheric structures. We use the Modern-Era Retrospective analysis for Research and Applications Version 2 (MERRA2), produced by NASA (Gelaro et al. 2017). This is a single-trajectory (i.e. not an ensemble) reanalysis product with data beginning in 1980, which exploits several observation types (Koster et al. 2016). It uses the GEOS-5 model with a spatial resolution of about 50 km in the horizontal and with 72 vertical levels, with the uppermost level at 0.01 hPa. The temporal resolution is 3 hours. With these reanalysis data, we create a realistic $\boldsymbol{\mu}^b$ and **B** to characterise the MVG to sample from. The elements in this process are detailed with the help of Figure 2 and Figure 1.

17
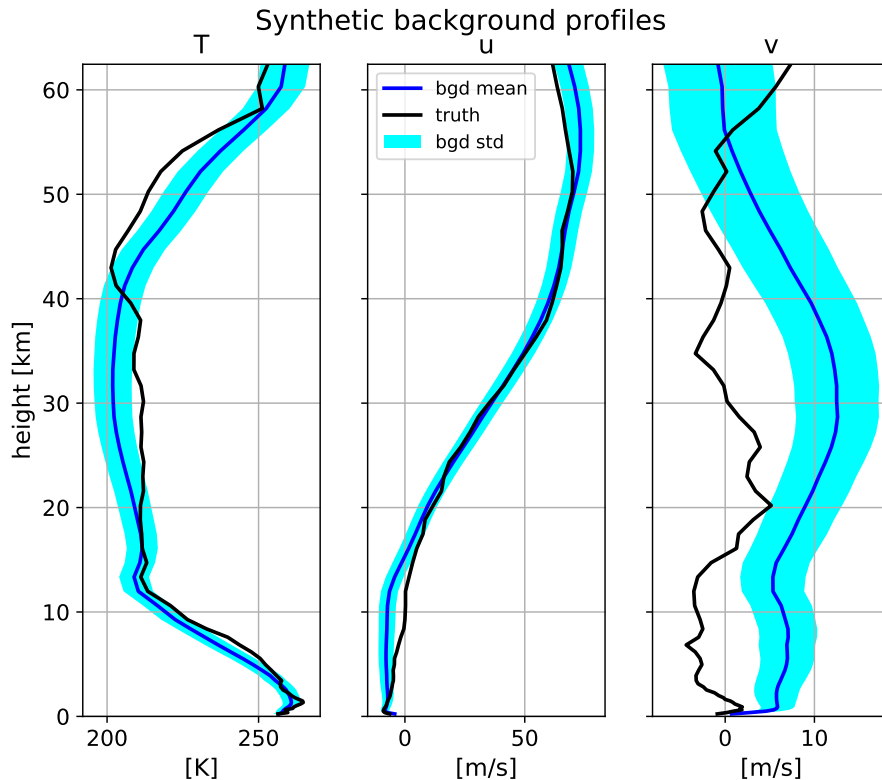
Synthetic background profiles

FIG. 2. Background mean (blue lines) plus/minus 1 standard deviation (shaded cyan area) for the state variables. The black lines represent the true profile used in the single-observation experiment.

1. We focus in the Northern Hemisphere autumn/winter period which is when the stratospheric winds are strong and have the greatest influence on the propagation of infrasound waves. This is also the period when the stratosphere-troposphere coupling is most prominent, and therefore a better representation of stratospheric dynamics can have the greatest impact on (tropospheric) numerical weather prediction (e.g., Charlton-Perez et al. 2013). We select 207 MERRA2 land grid points over Finland and northeast Norway (9 longitude by 23 latitude points). We consider five days in November 2010 (17th to 21st, inclusive). To represent common stratospheric conditions (e.g., a strong westerly polar vortex), we make sure to select days without sudden stratospheric warming events. We extract data for five times a day (every three hours, 9h–21h UTC).

2. The previous step provides a total of 5175 vertical profile samples of the $\{u(z), v(z), T(z)\}$ variables. We extract the lowermost $N_z = 60$ model levels, spanning from the surface to about

62 km altitude. We use these profiles to compute a correlation matrix **C**, shown in Figure 1. This is a block matrix where each block corresponds to a different physical variable, at all levels. The blocks are highlighted with black lines.

3. We use the statistics coming from the profiles in (1) to prescribe a MVG distribution from which to sample. We do not use these profiles directly since they do not follow an MVG distribution –there is no particular reason for them to do this, since these are columns coming from different locations at different times of a single-trajectory reanalysis. This is different from having, for example, a 12-hour ensemble forecast for a given location. Hence, we construct the distribution as follows. We choose one of the profiles from (a) as $\boldsymbol{\mu}^b$. Mimicking a realistic scenario, we prescribe background standard deviations that increase with altitude. For T, these standard deviations are set to vary linearly from 2 K to 8 K, while for both u and v they are set to vary linearly from 1.5 m/s to 6 m/s. These over-simplified values were partially inspired in results from (Amezcua and Barton 2021). With these elements, we construct the background covariance matrix **B** from:

$$\mathbf{B} = \Sigma^{\frac{1}{2}} \mathbf{C} \left(\Sigma^{\frac{1}{2}} \right)^{\top}, \tag{25}$$

where $\Sigma^{\frac{1}{2}}$ is a diagonal matrix containing the background standard deviations. Note that the climatological way in which we generate **B** is akin to an ensemble Optimal Interpolation (e.g. Sakov and Sandery 2015) implementation.

4. Having prescribed an MVG distribution, we generate a random sample ensemble with $N_e + N_{\text{true}} = 3200$ members. We take one member as the reference truth for the single-observation experiment, which is displayed with black line in the different panels of Figure 2, one for each variable. The thick blue lines represent $\boldsymbol{\mu}^b$, and the cyan shaded areas indicate one standard deviation it. Individual ensemble members are not shown to avoid cluttering.

b. Localization matrix design

Since we have $N_{\text{var}} = 3$ physical variables (each with $N_z = 60$ levels), both **B** and **L** are block matrices with $N_{\text{var}}^2 = 9$ blocks. As explained in Lee et al. (2022), different localization length-scales can be used for different blocks of **L**. For simplicity, in our implementation each block follows a
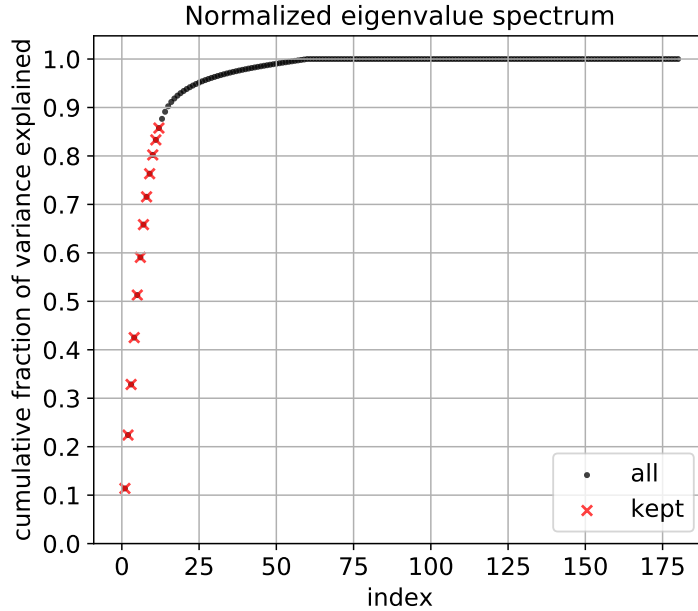
FIG. 3. Normalized eigenvalue spectrum of **L**. Only 12 (red dots) out of 180 (black) eigenvalues are kept.

Gaspari–Cohn decay from the main diagonal with a fixed length scale $l_z = 0.5$ km for all altitudes. Even though we did not perform a proper sensitivity analysis to the localization length scale, some brief experiments suggested that this length-scale yielded the best DA performance results from the list $l_z = \{0.5, 1.0, 1.5, 2.0, 2.5\}$ km.

Figures 3 and 4 illustrate the construction of the low-rank approximation to **L**. Figure 3 is a Scree plot, where the cumulative sum of its 180 normalized eigenvalues is shown. We keep the first 12 (shown in red), which account for approximately 86% of the variability. In figure 4 we show reconstructions with an increasing number of eigenvalues kept $N_\lambda = \{3, 6, 9, 12, 15\}$. In the bottom right panel we show the exact **L**. The case with $N_\lambda = 12$ eigenvalues kept renders a very good approximation with no visible spurious features like the ones observed in the panels of the first row.

c. Application of the observation operator

Mapping state variables into observation space is required for two purposes. The first is to generate synthetic observations (constructed from $\mathbf{x}^{\text{true}}$) to be assimilated. The second is to get $\mathbf{Y}^b$ from $\mathbf{X}^b$, as part of the analysis step of the ETKF algorithm.

20

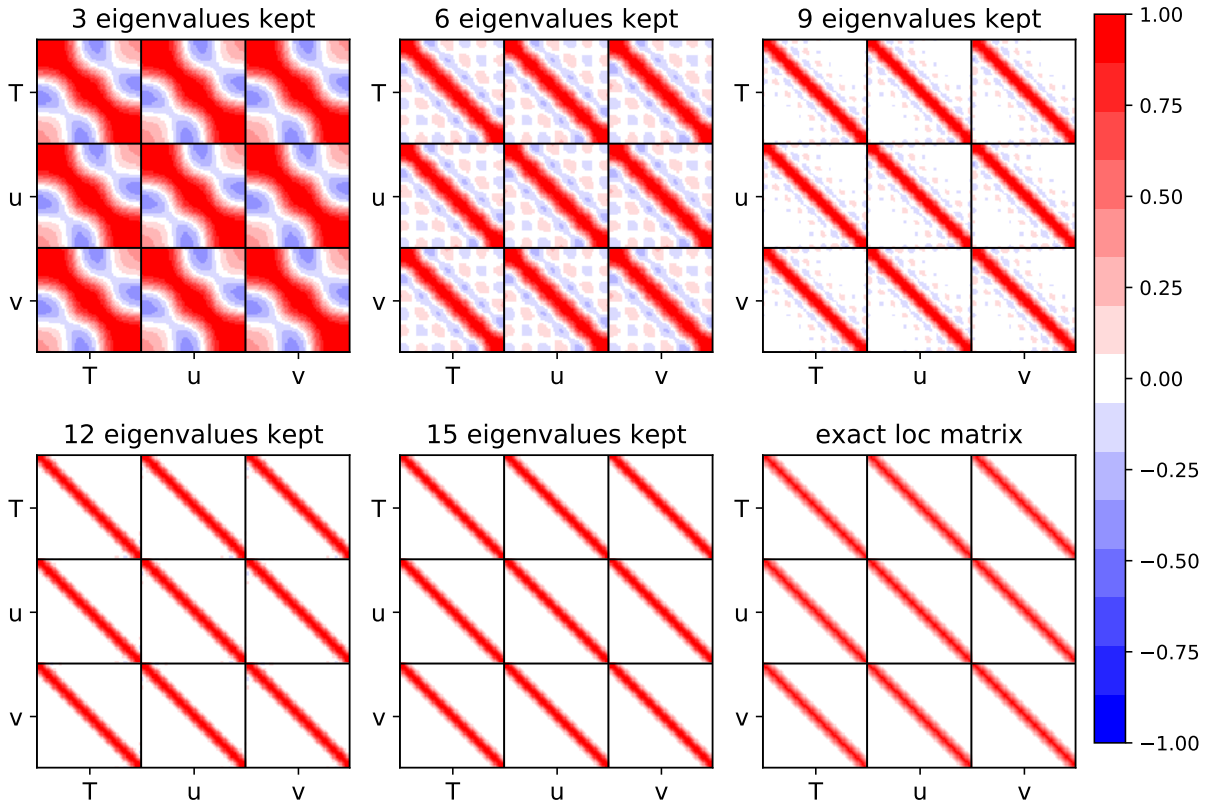Comparison of approximate and exact localization matrices

FIG. 4. Exact **L** (left) and its low-rank approximation **L̃**, based on 12 eigenvectors (right).

We use the InfraGA ray-tracing code to model infrasound wave propagation. This code is provided under open access by Los Alamos National Laboratory at `https://github.com/LANL-Seismoacoustics/infraGA` (e.g., Blom and Waxler (2021, 2017); last accessed 04 March 2024). Based on the ray-tracing simulations with mid-point reflection, we perform an eigenray search in 3D using the stratified moving-medium Cartesian propagation mode of InfraGA (which applies a horizontally homogeneous atmospheric profile). As further elaborated in Blixt et al. (2019) and Vera Rodriguez et al. (2020), the propagation for this source-receiver setup involves inserting a reflection point half-way along the great-circle, i.e. 90 km from the explosion site. It is assumed that the corresponding dominant reflection occurs at the single altitude Z_{max}. This is fixed at $Z_{max} = 38$ km based on previous works (Blixt et al. 2019; Amezcua et al. 2020; Amezcua and Barton 2021). In the search for an eigenray, we scan inclination angles from $10°$ to $80°$, and azimuth angle deviations of $\pm 0.5°$. After running InfraGA with these settings, 31 out of
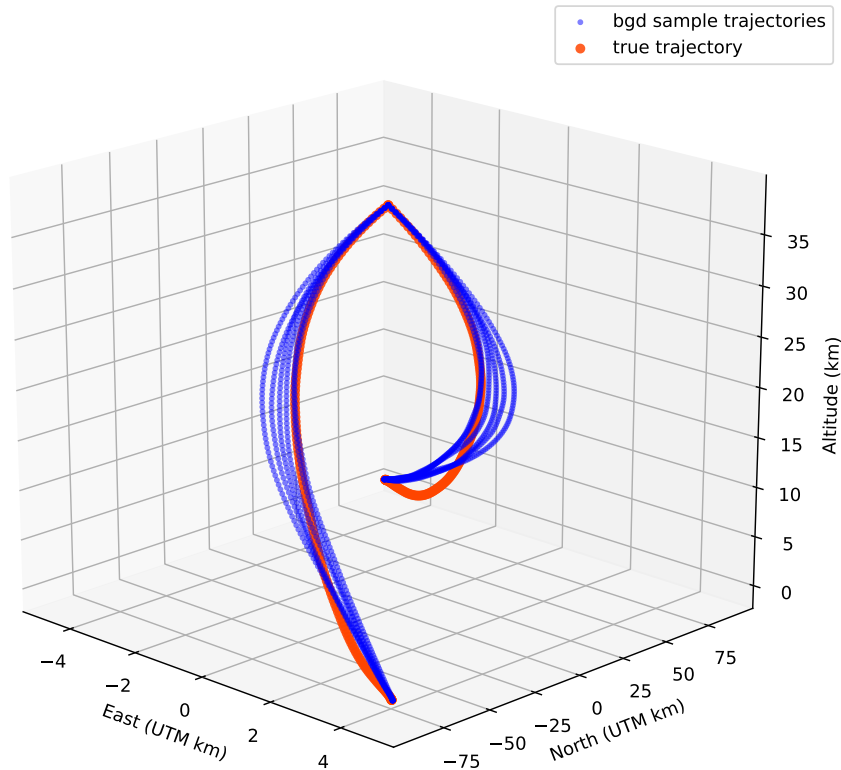
Fɪɢ. 5. True (dark orange) and background (blue) trajectories coming from Hukkakero (bottom) to ARCES (top), reflected at Z_{max} = 38 km.

the 3200 available profiles (0.96%) did not return the predicted observations. We examined the vertical profiles in which the search for an eigenray failed, and found that they share an average head wind of $v \geq 10 \; m/s$. These conditions do not allow for the physical existence of the desired eigenray. We remove these profiles from the background ensemble. This removal does not change the background statistics beyond the second decimal, with the largest changes in **v**.

Figure 5 shows a subset of the 3169 successful eigenray examples found. In this example, we show 5 background trajectories (blue lines), and one corresponding to the synthetic truth (red line) in the single-observation experiment. The axes correspond to the 3-D spatial dimensions, where the propagation occurs from Hukkakero to ARCES, which is almost due North. We can appreciate the complexity of the trajectories which was lost in the past when we approximated them as simple triangles when using the geometric model.

22

For our experiments, we randomly separate our 3169 profiles into two groups: $N_e = 2500$ members for $\mathbf{X}^b$, and $N_{\text{true}} = 669$ truths to assess the DA performance. Figure 6 shows the resulting observation space values for the 2500 predicted observations. The first row shows the marginal distributions of the observed variables (blue histograms, with vertical blue line representing the sample mean). We use the Shapiro-Wilk test (Shapiro and Wilk 1965) with significance $\alpha = 0.05$ to asses the Gaussianity of these distributions. Both δ_{ba} and t_{travel} can be considered Gaussian, while t_{trace} fails the test ($p \sim 10^{-6}$), mostly due to its asymmetry (left skewness). For comparison, we draw the closest Gaussian pdf (one with the same mean and standard deviation) in each panel. The second row shows two-dimensional cross-sections of the three-dimensional observation space. The aim of DA is hence to transform the cloud of predicted observations (blue dots) into a cloud closer to the real observation (dark orange star), and then to map this change into the original physical atmospheric variables. For reference, the means of the observed variables are shown in blue lines.
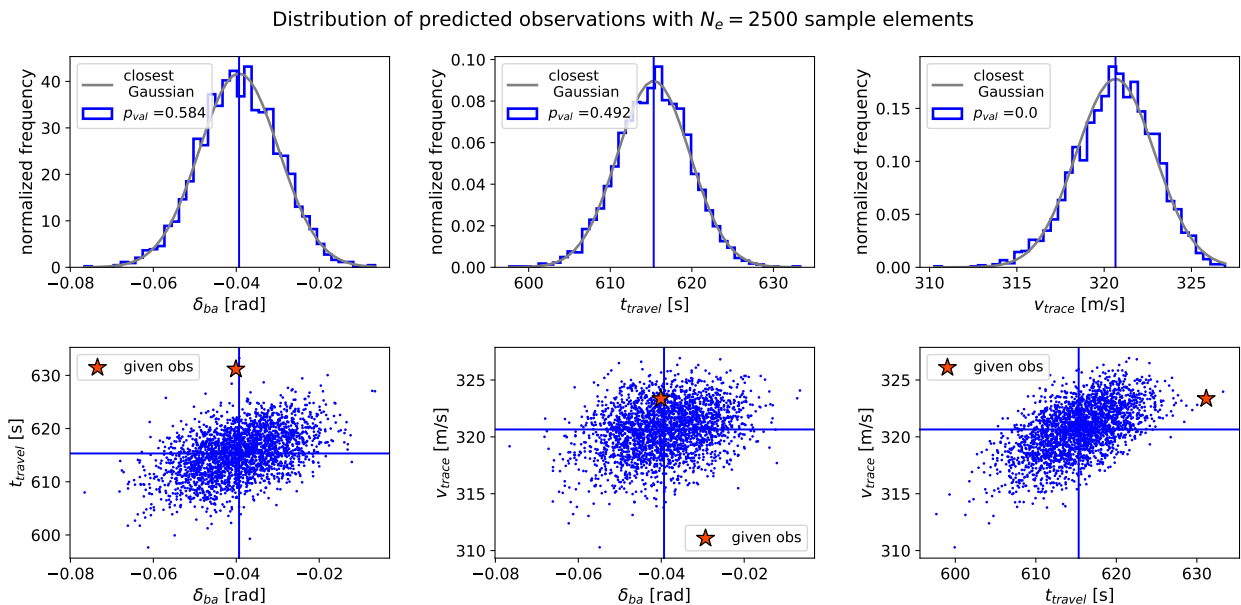

FIG. 6. Resulting predicted values in observation space for each background ensemble member (blue), and the true observation (dark orange). The first row assesses marginal Gaussianity for the observed quantities.

Finally, we need to prescribe observation error parameters. We consider $\boldsymbol{\eta}$ to be uncorrelated. Hence, $\mathbf{R}$ has the following values:

$$\mathbf{R} = \mathrm{diag}\left[\left(\frac{0.1\pi}{180}\,[\mathrm{rad}]\right)^2,\ (1\,[\mathrm{s}])^2,\ \left(0.5\left[\frac{\mathrm{m}}{\mathrm{s}}\right]\right)^2\right]. \tag{26}$$

Namely, the standard deviation of the observational errors are: $0.1°$ for δ_{ba}, 1 second for t_{travel}, and 0.5 m/s for v_{trace}, which agree with previous works. The observational error is quite low with respect to the spread of predicted observations, so we can expect large changes from $\mathbf{Y}^b$ to $\mathbf{Y}^a$ in the assimilation.

5. Applying METKF to synthetic infrasound data: Results

a. Evaluation metrics

For convenience, we present key quantities commonly used to evaluate the performance of a DA system. The analysis increment $\mathbf{d}^{ab} \in \mathbb{R}^{N_x}$ is defined as:

$$\mathbf{d}^{ab} = \bar{\mathbf{x}}^a - \bar{\mathbf{x}}^b, \tag{27}$$

i.e. this is the difference between the analysis and background means, which quantifies the effect observations have on the DA process. Hence, when there is no increment, there is no observational impact.

The OSSE allows for a direct comparison of our results to a synthetic reference truth. A common metric is the root-mean-square error of both $\bar{\mathbf{x}}^b$ and $\bar{\mathbf{x}}^a$ with respect to the synthetic truth:

$$\mathrm{RMSE}^i = \sqrt{\overline{(\bar{\mathbf{x}}^i - \mathbf{x}^{\mathrm{true}})^2}}, \quad i = \{b, a\} \tag{28}$$

where the square is evaluated element-wise, and the average (denoted by the overbar) can be taken in space, over time (when several DA cycles are performed sequentially), or over different experiment realizations. In cases with state variables of different physical units, it is common to compute this metric separately or use an energy norm, see, e.g., Kalnay (2003). In our numerical experiments, we compute these metrics for each physical variable at each vertical level. Hence, both expressions in Eq (28) are vectors with N_x elements.
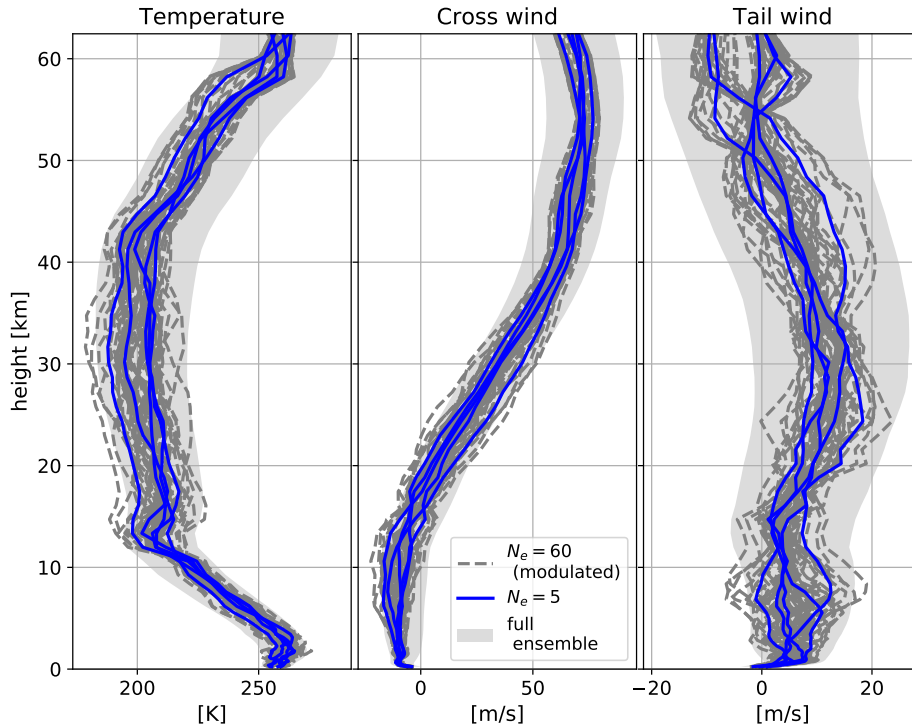
24

FIG. 7. Illustration of two background ensembles. Each panel corresponds to vertical profiles of different physical variables. The dark blue lines to the small ensemble ($N_e = 5$), and the dashed grey lines to the modulated ensemble ($N_{mod} = 60$). The gray shaded area corresponds to the mean plus minus three standard deviations coming from the full background ensemble ($N_e = 2500$).

b. Choices for background ensemble

We apply (M)ETKF in four cases:

1. A large $\mathbf{X}^b$ with $N_e = 2500$. This produces a matrix $\mathbf{P}^b$ indistinguishable from the one prescribed in Subsection a.

2. A small $\mathbf{X}^b$ with $N_e = 5$ which is useful to illustrate errors coming from small samples.

3. A modulated $\mathbf{X}^b$ with $N_{mod} = 60$. This comes from modulating the small ensemble in point (2) with the $N_\lambda = 12$ leading eigenvectors of $\mathbf{L}$.

4. A raw ensemble with $N_e = 60$ for comparison with point (3).

Figures 7 and 8 illustrate aspects of the modulated ensemble. Figure 7 shows the profiles for $T(z)$ (left), $u(z)$ (middle), and $v(z)$ (right). The gray shaded area represents the full background

25

Accepted for publication in *Monthly Weather Review.* DOI 10.1175/MWR-D-23-0186.1.

Unauthenticated | Downloaded 05/24/24 06:38 AM UTC

ensemble ($N_e = 2500$). To avoid cluttering we do not show the individual ensemble members, but instead a region corresponding to the mean plus minus 3 standard deviations for each vertical level. The dark blue lines correspond to the randomly chosen small ensemble with $N_e = 5$. The gray dashed lines correspond to the modulated ensemble with $N_e = 60$. By construction, the mean and standard deviation of the small and the modulated ensemble coincide. However, the range of values covered by the modulated ensemble is larger.
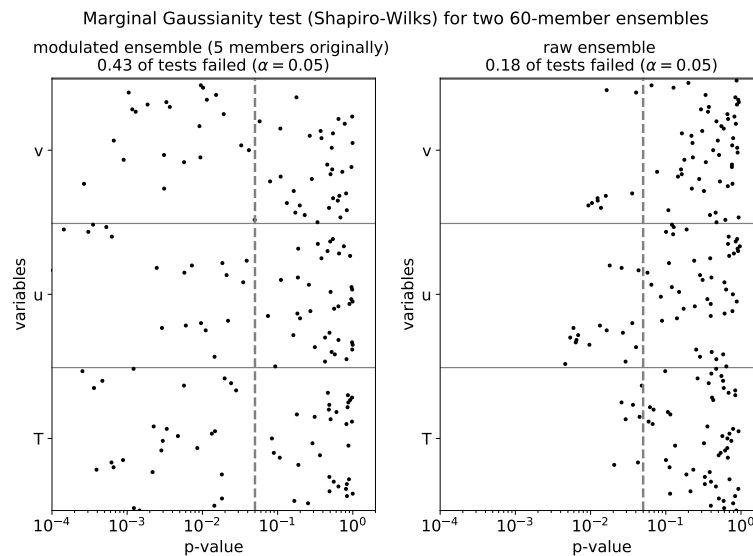
FIG. 8. Marginal Gaussianity for the state variables in two ensembles with 60 members. The left panel corresponds to a modulated ensemble, while the right column to a raw ensemble

.

We generated additional modulated ensembles where we kept fewer eigenvalues of **L**, in particular $N_\lambda = \{3, 6, 9\}$. However, they resulted in profiles with pronounced wave-like vertical variations related to the spurious features discussed in Figure 4. These were therefore discarded and not shown for brevity.

Before moving on, it is interesting to ponder whether the modulation process changes the statistical characteristics of the ensemble. To answer this question, we compare the modulated ensemble $N_{mod} = 60$ with a raw ensemble of the same size. For each variable at each vertical level, we perform a Shapiro-Wilk test with significance $\alpha = 0.05$ for marginal Gaussianity. The p-values for both ensembles are shown in Figure 8 (note the logarithmic scale for the x-axis).

Let us start with the right panel. If all the variables were independent, the expected fraction of failed local tests with $\alpha = 0.05$ would be 0.05. Due to the covariance structure, this increases to 0.18. We can take this number as our basis to pass or fail a field significance test (see e.g. Livezey and Chen (1983) for the distinction between local and field significance). In the left panel we have the results for the modulated ensemble. In this case, the fraction of failed local tests increases to 0.46. Furthermore, the p-values reach lower values than in the case of the raw ensembles. This suggests that the modulated ensemble has lost Gaussianity with respect to the raw ensemble. Our empirical result agrees with the theoretical loss of Gaussianity in modulated ensembles demonstrated in Chan (2023).
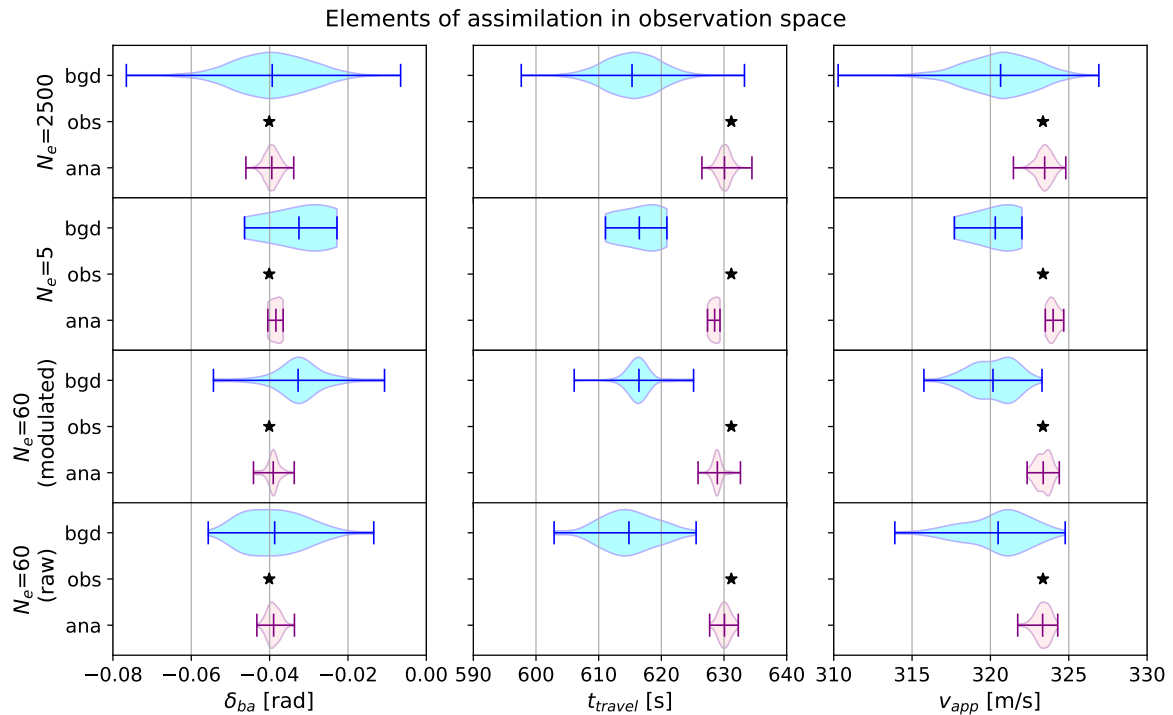

FIG. 9. Results in observation space of the application of (M)ETKF in the single observation experiment. Each row corresponds to a different ensemble size, and each column a different observed quantity. The violin plots represent the background (cyan/blue) and analysis (pink/purple), where the ticks correspond to minimum, mean and maximum value. The black star corresponds to the given observation.

27

c. Results

First we show the results for the single-observation experiment. Figure 9 displays the DA elements in observation space, where columns correspond to different observed quantities and rows correspond to a different background ensemble. In each panel, the background (blue) and analysis (purple) elements are visualized with violin plots (Hintze and Nelson 1998), while the given observation is shown with a black star. In observation space, the assimilation is successful in all cases. The analysis ensembles are located closer to the given observation than the background ensembles, and they have considerably reduced spread.
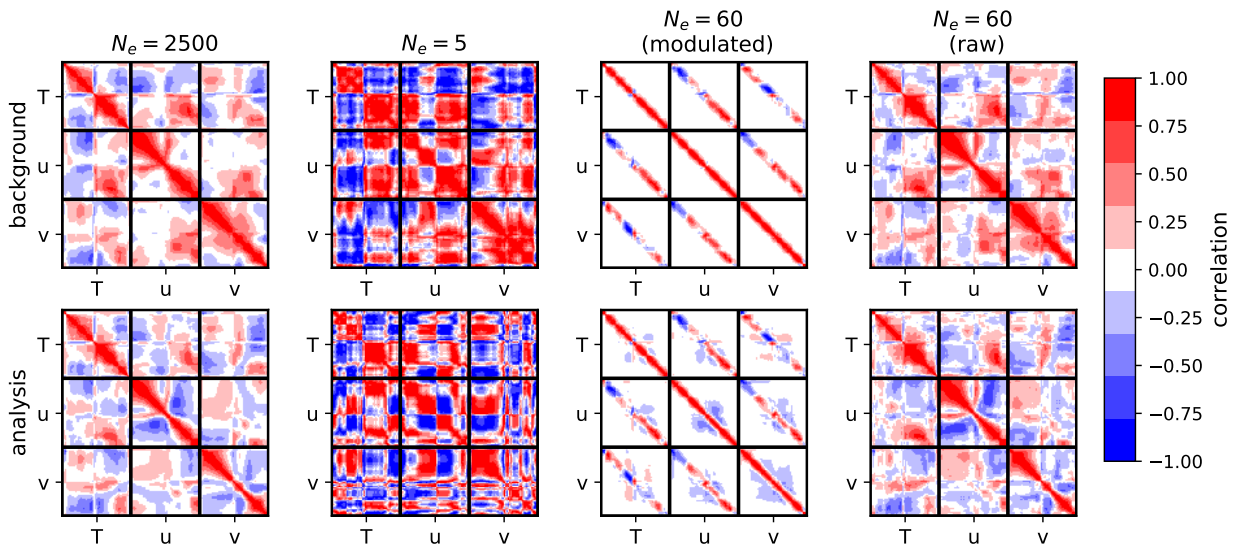

FIG. 10. Background (top) and analysis (bottom) correlation matrices involved in the DA experiments with infrasound observations.

Figure 10 depicts the sample background (top) and analysis (bottom) correlation matrices, which we label $\mathbf{C}^b$ and $\mathbf{C}^a$ respectively. We display these instead of the covariance matrices to avoid masking of the structures by the different magnitudes of the variables. Once more, we add thick black lines to denote the 9 block matrices they are made of. In the upper left, we have $\mathbf{C}^b$ generated by $N_e = 2500$. This is indistinguishable from the prescribed $\mathbf{C}$. The corresponding $\mathbf{C}^a$ has structures similar to those found in $\mathbf{C}^b$, but some changes are noticeable. For $N_e = 5$, the corresponding $\mathbf{C}^b$ is visibly contaminated by sampling noise, as expected. The corresponding $\mathbf{C}^a$ is therefore also contaminated by direct sampling noise, but also by spurious changes coming from the calculation of an erroneous gain. In the modulated case $\mathbf{C}^b$ is reduced to a matrix with thin

28

strips around the main diagonals of the 9 blocks, as expected. This structure is changed when computing $\mathbf{C}^a$, with information spreading to the previously localised areas. For $N_e = 60$ members, $\mathbf{C}^b$ is considerably similar to the one coming from $N_e = 2500$, although some sampling noise is still visible for off-diagonal elements in each one of the blocks.
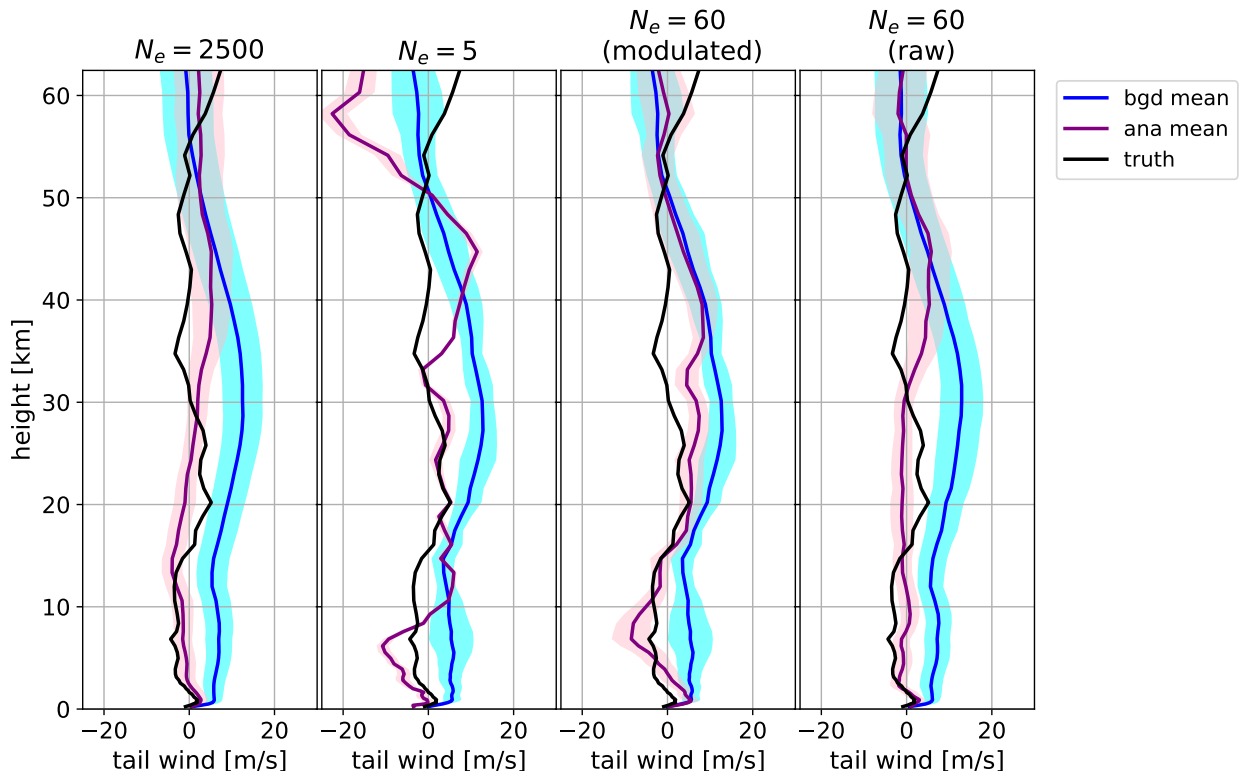

FIG. 11. Model space results for the single-observation experiments. Vertical profiles of $v(z)$ are shown for different ensemble sizes. Dark lines correspond to the reference truth (black), background mean (dark blue), and analysis mean (dark purple). Light shaded areas correspond to plus/minus 1 standard deviation (cyan for background, pink for analysis).

Figure 11 displays the DA elements in the model space. For brevity, we only display the results for the tail wind $v(z)$, which was the most benefited variable by the DA process (the results for the other two variables are qualitatively similar, but less pronounced). Each column corresponds to a different ensemble size. Each panel contains several lines: the true profiles are shown with thick black lines, the background profiles are shown in blue (dark blue for the mean, shaded cyan region for plus/minus 1 standard deviation), and the analysis profiles are shown in purple (dark purple the mean, shaded pink region for plus/minus 1 standard deviation).

29

The assimilation is effective for $N_e = 2500$ members, with $\bar{\mathbf{x}}^a$ being much closer to $\mathbf{x}^{true}$ than $\bar{\mathbf{x}}^b$. Furthermore, there is a reduction in the spread from $\mathbf{X}^b$ to $\mathbf{X}^a$. The case $N_e = 5$ shows different results. In this case, $\bar{\mathbf{x}}^a$ is in general further from $\mathbf{x}^{true}$ than $\bar{\mathbf{x}}^b$. We see pronounced errors in levels above $z = 35$ km and below $z = 10$ km. This happens despite the fact that the assimilation seems successful in observation space for the same ensemble size (Figure 9). This is because the spurious correlation structures render the projection into model space defective. Moreover, the analysis ensemble spread is too small (as expected from the small sample size), rendering an overconfident and inaccurate ensemble. The third column shows the modulated ensemble. The estimation is not as accurate as with $N_e = 2500$, but it avoids the incorrect update that occurred for $N_e = 5$. In general, $\bar{\mathbf{x}}^a$ is closer to $\mathbf{x}^{true}$ than $\bar{\mathbf{x}}^b$. The incorrect updates in the uppermost levels (which happened with $N_e = 5$) are no longer a problem when using the modulated ensemble. Using a raw ensemble with $N_e = 60$ yields very similar results to the case with $N_e = 2500$, which is not unexpected since this is a relatively small problem ($N_x = 60$).

After the visual examination of the results in Figure 11, we perform a more systematic quantification of the DA performance. For this purpose, we repeat the DA experiment for the $N_{\text{true}} = 669$ truths, each with a respective set of observations. These are generated with independent realizations of $\boldsymbol{\eta}$.

In Figure 12 we compute both RMSE^b (blue line) and RMSE^a (purple line) as prescribed in Eq (28) for each variable and vertical level separately. The top row shows the case for $N_e = 2500$. Here, the assimilation always improves the estimation, which is clear since the purple line is always to the left of the blue line. The largest error reduction occurs for both the wind components $u(z)$ and $v(z)$, with an error reduction close to 2 m/s around $z = 30$ km for both variables.

The second row corresponds to $N_e = 5$ and the third row to $N_{\text{mod}} = 60$. Both ensembles share the same $\bar{\mathbf{x}}^b$, and hence they have the same RMSE^b. As expected, the small ensemble ($N_e = 5$) yields $\text{RMSE}^a > \text{RMSE}^b$ for most vertical levels. This implies that assimilating the observation considerably damages the estimation due to the erroneous increments coming from sampling errors. These spurious errors can reach 3 K for $T(z)$ and 4 m/s for $u(z)$. The use of the modulated ensemble is effective in eliminating these errors for the three variables. For $T(z)$, $\text{RMSE}^a \approx \text{RMSE}^b$, but at least there are no negative consequences from the assimilation. For both $u(z)$ and $v(z)$, there is noticeable error reduction due to the assimilation in this modulated case.
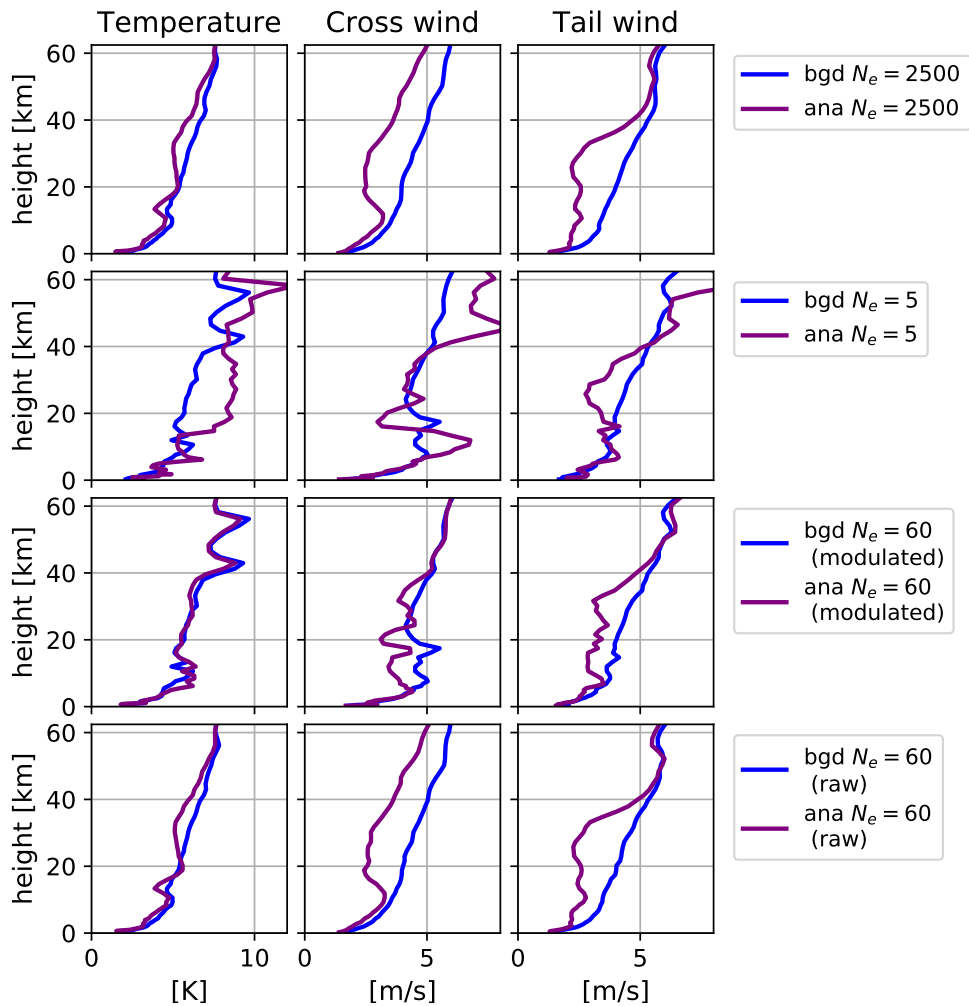
Fɪɢ. 12. Evaluation of the performance of the (M)ETKF in terms of RMSE of $\bar{\mathbf{x}}^b$ (blue line) and $\bar{\mathbf{x}}^b$ (purple lines) with respect to $\mathbf{x}_{\text{true}}$ for different ensemble sizes (rows) and variables (columns).

The last row shows the case with $N_e = 60$ members, which has RMSE values very similar to those coming from $N_e = 2500$ members. The performance of the modulated ensemble with $N_{mod} = 60$ members is considerably better than that with $N_e = 5$ members, but slightly worse than the raw ensemble with $N_e = 60$ members.

d. The impact of assimilating different observations

When designing new observation systems and instruments, or devising new uses for existing ones, one needs to determine the impact that the assimilation of these observations have on the state variables. This is often referred to as the forecast sensitivity to observations (FSOI) (e.g.,

Kotsuki et al. 2019). In its simplest form, FSOI can be performed with data denial experiments, which are straightforward in an OSSE setting. Given the same background values, different subsets of observations are assimilated independently to generate different analysis values.

For these experiments we use $N_e = 2500$ members, since we want information without sampling errors. We assimilate four subsets of observations:

$$\{\{\delta_{ba}\}, \{t_{\text{travel}}\}, \{v_{\text{trace}}\}, \{\delta_{ba}, t_{\text{travel}}, v_{\text{trace}}\}\}. \tag{29}$$

For each subset, we perform $N_{true} = 699$ experiments. For each experiment, we compute the analysis increment $\mathbf{d}^{ba}$ as defined in Eq (27). After all experiments, we have $N_{true} = 699$ vectors $\mathbf{d}^{ba}$. For each of the N_x components of the $\mathbf{d}^{ba}$ we compute the following percentiles: $\{10, 25, 50, 75, 90\}$. The results are displayed in Figure 13 as vertical profiles $T(z)$ (top row), $u(z)$ (middle row), $v(z)$ (bottom row). The columns correspond to the groups mentioned in Eq (29). When an observation has little or no impact for a particular variable at a particular level, percentiles are close to zero (the analysis is almost the same as the background). On the other hand, if there are big differences between the quantiles, this means that the observation has a large impact and leads to large analysis increments. We note the following:

1. For all variables, $\mathbf{d}^{ab}$ can be positive and negative. The 50th percentile for all panels is close to the zero line, and the $25^{\text{th}} - 75^{\text{th}}$ and $10^{\text{th}} - 90^{\text{th}}$ percentiles are qualitatively symmetric with respect to each other.

2. For $u(z)$, the most informative observation is δ_{ba}. It provides non-zero analysis increments throughout the whole atmospheric column, but particularly in the region from 15 to 50 km. The role of this observation is consistent with the role suggested by the Blixt geometric model (Blixt et al. 2019) and the findings of Amezcua et al. (2020). Both t_{travel} and v_{trace} provide some information above 20 km. The increments coming from t_{travel} increase with height, while the increments coming from v_{trace} peak at around 40 km, just above the reflection height.

3. For $v(z)$, the most informative observation is t_{travel}. This observation has a large impact from the surface to around 50 km high, peaking around 30 km. The next observation is δ_{ba}, with smaller increments peaking at around 30 km and decreasing both towards the surface and
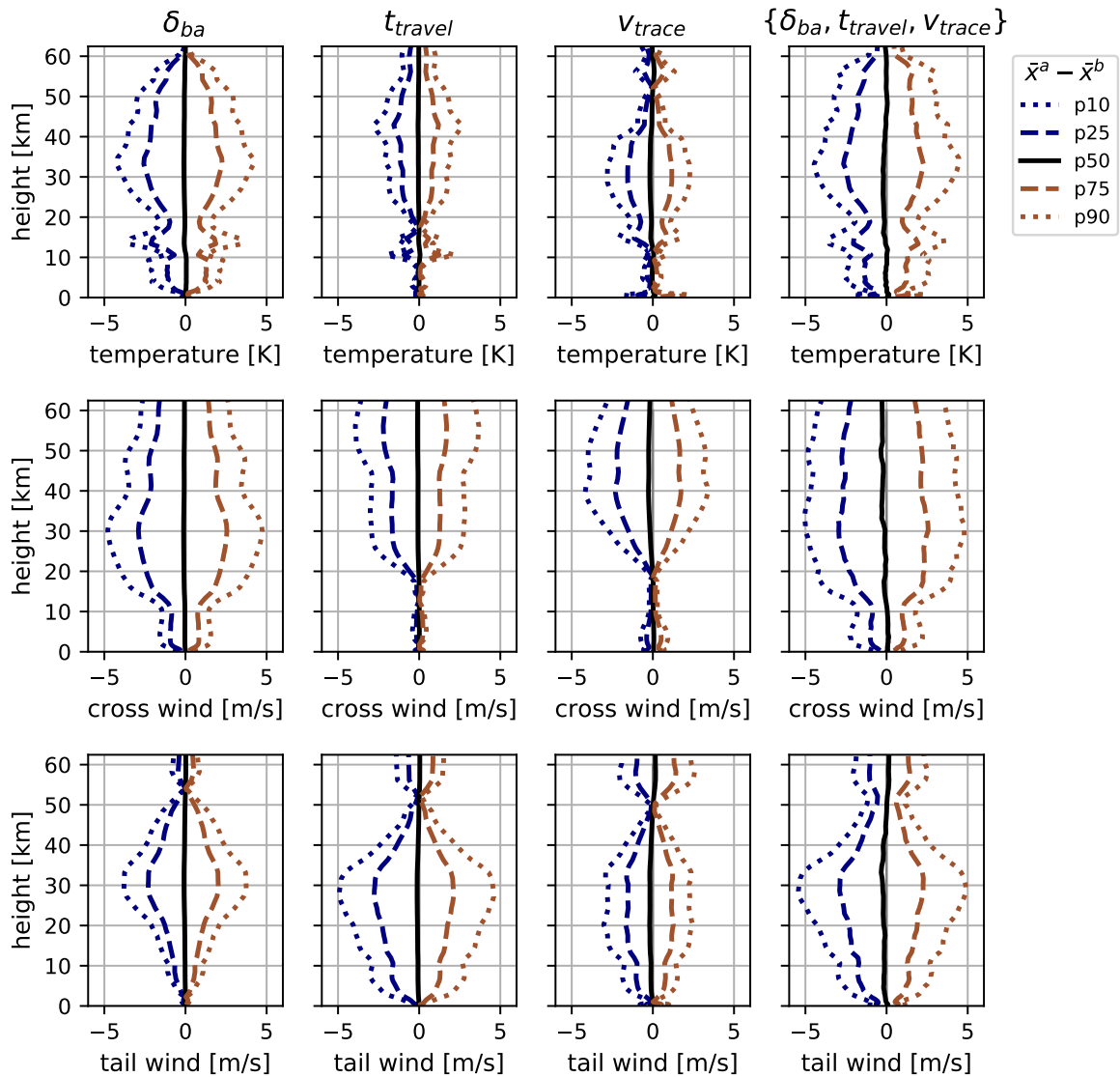
32

FIG. 13. Analysis impact from assimilating different observations (columns): only δ_{ba}, only t_{travel}, only v_{trace}, and three at the same time. The analysis increments are collected over 669 experiments. Each row shows a variable (T in the top, u in the middle, v in the bottom). For each panel, we show the percentiles of the increment.

towards the top of the column. Assimilating v_{trace} leads to moderate increments throughout the column, except for a region around 50 km.

6. Summary and discussion

The objective of this paper was to implement an ensemble Kalman filter to assimilate infrasound-related observations in order to update an atmospheric vertical profile. In this method, the quality

33

of the sampling estimators for mean and covariance is crucial, and these estimators often come from limited-size ensembles. Localization is a method often used in the EnKF community to reduce sampling errors. Infrasound-related observations contain spatially-integrated information, and can be difficult to localize. To achieve this, we borrowed a technique from the satellite DA community: the modulated ensemble Kalman filter.

We performed state estimation experiments to update three height-dependent fields: cross wind, tail wind, and temperature. We used 4 ensemble sizes: a large ensemble with $N_e = 2500$, a small ensemble with $N_e = 5$, a modulated ensemble with $N_e = 60$, and a (raw) ensemble with $N_e = 60$ (to compare against the modulated ensemble). The assimilation process in the three-dimensional observation space was successful in all cases. The mapping back to model space proved to be the main challenge. With a 2500-member ensemble, we were able to reduce the RMSE of the analysis mean in the middle atmosphere (around 30 km) from that of the background mean by about 2 m/s for the horizontal winds and about 1 K for the temperature. Using a 5-member ensemble led to poor results (as expected), with the RMSE of the analysis mean being larger than that of the background. The use of a 60-member modulated ensemble (5 original members modulated by the leading 12 eigenvectors of the designed localization matrix) helped eliminate spurious increments. The RMSE of the analysis mean was seldom larger than that of the background, and we even had some reductions (although smaller than for the large ensemble). A 60-member raw ensemble showed results similar to that of the 2500-member ensemble, and slightly better than that of the modulated ensemble. With the large-ensemble setting, we also performed sensitivity analysis experiments where we related the assimilation of each of the three observations by themselves to the impacts they have in the vertical profiles at different levels.

This study demonstrates the possibility of assimilating infrasound observations using ensemble DA with a full ray-propagation model as the observation operator. Furthermore, we have shown how satellite DA techniques can be readily applied to ensure appropriate localization and therefore make small ensembles useful. While we have shown some quantitative evidence of the improvement in the estimation of quantities in vertical columns, it is important to remember that the ultimate aim is to assimilate novel observations (in our case the infrasound-related quantities) in tandem with the traditionally available observations like radiosondes (see e.g. Francis et al. (2023) for a discussion on the role of radiosondes as anchor observations in remote sensing).

34

This study generates some questions we are tackling in our current work. These are:

1. What does it mean for a ray-tracing routine to fail and what information does this failure give to the DA process? As we mentioned 0.96% of the background ensemble members did not generate a return when fed to InfraGA. If there were no errors in the observation operator, this would mean that the likelihood of the observation occurring for a given background profile is null, discarding the corresponding ensemble member (due to strong head winds like in our case). In a particle filter setting (e.g van Leeuwen 2009), this would mean a zero posterior weight. In an EnKF setting, this is more complicated and how to include this information properly needs to be explored. It is important to mention that we did not change Z_{max} when running InfraGA. For instance, Blixt et al. (2019) performed ray-tracing with different reflection heights and did not experience any failure for the vertical profiles they considered.

2. This study considered a prescribed Z_{max}. A more ambitious aim is to estimate this unknown parameter within the DA process, i.e. to perform experiments with a joint estimation of state and parameters. Furthermore, in reality there are partial reflections from multiple heights. These reflection heights could be estimated simultaneously. The interaction of these parameters with the observations, however, is complex and requires detailed analysis.

3. The assimilation process in a range-dependent (horizontally varying) regime is a larger goal, especially when the transmission occurs over longer horizontal distances. Amezcua and Barton (2021) performed this type of assimilation with the geometric model, but a range-dependent ray-tracing routine could be applied. This is still computationally feasible. Applying more exact wave-propagation routines like the solution of parabolic partial differential equations is, for the moment, out of our scope. The reader is referred to Brissaud et al. (2022) for a discussion of these types of models and their emulation using machine learning techniques for computational efficiency.

4. In this study, the application of the observation operator yielded observed quantities which behaved close to Gaussian, with trace velocity being the quantity with the largest departure from this distribution. The modulation process within METKF yielded state variables which lost marginal Gaussianity with respect to the original ensemble. This did not seem to have harmful effects in the assimilation process. However, it is worth exploring these effects in the

35

range-dependent regime and assessing whether extra steps (e.g. Gaussian anamorphosis) are needed.

This study has underlined the possibility of exploiting under-used datasets. Middle- and upper-atmospheric observations are considerably sparser –both spatially and temporally– than those covering the lower atmosphere. Therefore, any existing observations, albeit indirect, are potentially useful in NWP.

There is a call for assimilating more wind-related observations in the middle atmosphere and it is notable that 3DVar assimilation experiments have also been performed for winds in the MLT using meteor radar multi-station datasets (Stober et al. 2021). Our long-term goal is to assimilate infrasound-related observations from natural sources like ocean-generated microbaroms. These processes are continuous in time and space. This provides a constant observational record, but also challenging aspects in the modelling and assimilation fronts.

Data availability statement. The MERRA2 re-analysis data are openly available. It can be retrieved from the following URL: `https://gmao.gsfc.nasa.gov/reanalysis/MERRA-2/`. The synthetic atmospheric profiles and their mappings into observation space can be found in Amezcua's GitHub repository `https://github.com/jamezcuae/infrasound`.

References

Amezcua, J., and Z. Barton, 2021: Assimilating atmospheric infrasound data to constrain atmospheric winds in a two-dimensional grid. *Quarterly Journal of the Royal Meteorological Society*, **147 (740)**, 3530–3554, https://doi.org/10.1002/qj.4141.

Amezcua, J., S. P. Näsholm, E. M. Blixt, and A. J. Charlton-Perez, 2020: Assimilation of atmospheric infrasound data to constrain tropospheric and stratospheric winds. *Quarterly Journal of the Royal Meteorological Society*, **146 (731)**, 2634–2653, https://doi.org/10.1002/qj.3809.

Anderson, J. L., 2001: An ensemble adjustment Kalman filter for data assimilation. *Monthly Weather Review*, **129 (12)**, 2884–2903, https://doi.org/10.1175/1520-0493(2001)129⟨2884:AEAKFF⟩2.0.CO;2.

Anderson, J. L., 2007: Exploring the need for localization in ensemble data assimilation using a hierarchical ensemble filter. *Physica D: Nonlinear Phenomena*, **230 (1)**, 99–111, https://doi.org/10.1016/j.physd.2006.02.011.

Anderson, J. L., 2012: Localization and sampling error correction in ensemble Kalman filter data assimilation. *Monthly Weather Review*, **140 (7)**, 2359–2371, https://doi.org/10.1175/MWR-D-11-00013.1.

Arnold, C. P., and C. H. Dey, 1986: Observing-systems simulation experiments: Past, present, and future. *Bulletin of the American Meteorological Society*, **67 (6)**, 687–695, https://doi.org/10.1175/1520-0477(1986)067⟨0687:OSSEPP⟩2.0.CO;2.

Arrowsmith, S. J., O. Marcillo, and D. P. Drob, 2013: A framework for estimating stratospheric wind speeds from unknown sources and application to the 2010 December 25 bolide. *Geophysical Journal International*, **195 (1)**, 491–503, https://doi.org/10.1093/gji/ggt228.

Asch, M., M. Bocquet, and M. Nodet, 2016: *Data Assimilation*. Society for Industrial and Applied Mathematics, Philadelphia, PA, https://doi.org/10.1137/1.9781611974546.

Assink, J., P. Smets, O. Marcillo, C. Weemstra, J.-M. Lalande, R. Waxler, and L. Evers, 2019: Advances in infrasonic remote sensing methods. *Infrasound Monitoring for Atmospheric Studies*, Springer, 605–632, https://doi.org/10.1007/978-3-319-75140-5_18.

37

Assink, J. D., A. L. Pichon, E. Blanc, M. Kallel, and L. Khemiri, 2014: Evaluation of wind and temperature profiles from ECMWF analysis on two hemispheres using volcanic infrasound. *Journal of Geophysical Research: Atmospheres*, **119 (14)**, 8659–8683, https://doi.org/10.1002/2014JD021632.

Assink, J. D., R. Waxler, W. G. Frazier, and J. Lonzaga, 2013: The estimation of upper atmospheric wind model updates from infrasound data. *Journal of Geophysical Research: Atmospheres*, **118 (19)**, 10–707, https://doi.org/10.1002/jgrd.50833.

Baldwin, M. P., and T. J. Dunkerton, 2001: Stratospheric harbingers of anomalous weather regimes. *Science*, **294 (5542)**, 581–584, https://doi.org/10.1126/science.106331.

Baldwin, M. P., D. B. Stephenson, D. W. J. Thompson, T. J. Dunkerton, A. J. Charlton, and A. O'Neill, 2003: Stratospheric memory and skill of extended-range weather forecasts. *Science*, **301 (5633)**, 636–640, https://doi.org/10.1126/science.1087143.

Bauer, P., A. Thorpe, and G. Brunet, 2015: The quiet revolution of numerical weather prediction. *Nature*, **525 (7567)**, 47–55, https://doi.org/10.1038/nature14956.

Bishop, C. H., B. J. Etherton, and S. J. Majumdar, 2001: Adaptive sampling with the ensemble transform Kalman filter. Part I: Theoretical aspects. *Monthly Weather Review*, **129 (3)**, 420–436, https://doi.org/10.1175/1520-0493(2001)129⟨0420:ASWTET⟩2.0.CO;2.

Bishop, C. H., and D. Hodyss, 2009: Ensemble covariances adaptively localized with ECO-RAP. Part 2: a strategy for the atmosphere. *Tellus A*, **61 (1)**, 97–111, https://doi.org/10.1111/j.1600-0870.2008.00372.x.

Bishop, C. H., 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. *Monthly Weather Review*, **145 (11)**, 4575–4592, https://doi.org/10.1175/MWR-D-17-0102.1.

Blanc, E., and Coauthors, 2018: Toward an improved representation of middle atmospheric dynamics thanks to the ARISE project. *Surveys in Geophysics*, **39 (2)**, 171–225, https://doi.org/10.1007/s10712-017-9444-0.

Blanc, E., and Coauthors, 2019: Middle atmosphere variability and model uncertainties as investigated in the framework of the ARISE project. *Infrasound Monitoring for Atmospheric Studies*, Springer, 845–887, https://doi.org/10.1007/978-3-319-75140-5_28.

Blixt, E. M., S. P. Näsholm, S. J. Gibbons, L. G. Evers, A. J. Charlton-Perez, Y. J. Orsolini, and T. Kværna, 2019: Estimating tropospheric and stratospheric winds using infrasound from explosions. *The Journal of the Acoustical Society of America*, **146 (2)**, 973–982, https://doi.org/10.1121/1.5120183.

Blom, P., and R. Waxler, 2017: Modeling and observations of an elevated, moving infrasonic source: Eigenray methods. *The Journal of the Acoustical Society of America*, **141 (4)**, 2681–2692, https://doi.org/10.1121/1.4980096.

Blom, P., and R. Waxler, 2021: Characteristics of thermospheric infrasound predicted using ray tracing and weakly non-linear waveform analyses. *The Journal of the Acoustical Society of America*, **149 (5)**, 3174–3188, https://doi.org/10.1121/10.0004949.

Blom, P. S., O. Marcillo, and S. J. Arrowsmith, 2015: Improved Bayesian infrasonic source localization for regional infrasound. *Geophysical Journal International*, **203 (3)**, 1682–1693, https://doi.org/10.1093/gji/ggv387.

Bluestein, H. B., F. H. Carr, and S. J. Goodman, 2022: Atmospheric observations of weather and climate. *Atmosphere-Ocean*, **60 (3-4)**, 149–187, https://doi.org/10.1080/07055900.2022.2082369.

Bonavita, M., and P. Laloyaux, 2020: Machine learning for model error inference and correction. *Journal of Advances in Modeling Earth Systems*, **12 (12)**, e2020MS002 232, https://doi.org/10.1029/2020MS002232.

Brissaud, Q., S. P. Näsholm, A. Turquet, and A. Le Pichon, 2022: Predicting infrasound transmission loss using deep learning. *Geophysical Journal International*, **232 (1)**, 274–286, https://doi.org/10.1093/gji/ggac307, URL https://doi.org/10.1093/gji/ggac307, https://academic.oup.com/gji/article-pdf/232/1/274/45967048/ggac307.pdf.

Campbell, W. F., C. H. Bishop, and D. Hodyss, 2010: Vertical covariance localization for satellite radiances in Ensemble Kalman filters. *Monthly Weather Review*, **138 (1)**, 282 – 290, https://doi.org/10.1175/2009MWR3017.1.

Chan, M.-Y., 2023: Improving ensemble data assimilation through probit-space ensemble size expansion for gaussian copulas (pese-gc). *EGUsphere*, **2023**, 1–26, https://doi.org/10.5194/egusphere-2023-2699, URL https://egusphere.copernicus.org/preprints/2023/egusphere-2023-2699/.

Charlton-Perez, A. J., and Coauthors, 2013: On the lack of stratospheric dynamical variability in low-top versions of the CMIP5 models. *Journal of Geophysical Research: Atmospheres*, **118 (6)**, 2494–2505, https://doi.org/10.1002/jgrd.50125.

Cugnet, D., A. de la Camara, F. Lott, C. Millet, and B. Ribstein, 2019: Non-orographic gravity waves: Representation in climate models and effects on infrasound. *Infrasound Monitoring for Atmospheric Studies: Challenges in Middle Atmosphere Dynamics and Societal Benefits*, A. Le Pichon, E. Blanc, and A. Hauchecorne, Eds., Springer, Cham, Switzerland, 827–844, https://doi.org/10.1007/978-3-319-75140-5_27.

Dee, D. P., and Coauthors, 2011: The ERA-Interim reanalysis: configuration and performance of the data assimilation system. *Quarterly Journal of the Royal Meteorological Society*, **137 (656)**, 553–597, https://doi.org/10.1002/qj.828.

Drob, D. P., and Coauthors, 2008: An empirical model of the earth's horizontal wind fields: Hwm07. *Journal of Geophysical Research*, **113 (A12)**, https://doi.org/10.1029/2008JA013668.

Evensen, G., 1994: Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics. *Journal of Geophysical Research: Oceans*, **99 (C5)**, 10 143–10 162, https://doi.org/10.1029/94JC00572.

Evensen, G., 2006: *Data Assimilation: The Ensemble Kalman Filter*. Springer-Verlag, Berlin, Heidelberg, https://doi.org/10.1007/978-3-642-03711-5.

Eyre, J. R., W. Bell, J. Cotton, S. J. English, M. Forsythe, S. B. Healy, and E. G. Pavelin, 2022: Assimilation of satellite data in numerical weather prediction. Part II: Recent years.

Quarterly Journal of the Royal Meteorological Society, **148 (743)**, 521–556, https://doi.org/10.1002/qj.4228.

Eyre, J. R., S. J. English, and M. Forsythe, 2020: Assimilation of satellite data in numerical weather prediction. Part I: The early years. *Quarterly Journal of the Royal Meteorological Society*, **146 (726)**, 49–68, https://doi.org/10.1002/qj.3654.

Farchi, A., and M. Bocquet, 2019: On the efficiency of covariance localisation of the ensemble kalman filter using augmented ensembles. *Frontiers in Applied Mathematics and Statistics*, **5**, https://doi.org/10.3389/fams.2019.00003.

Francis, D. J., A. M. Fowler, A. S. Lawless, J. Eyre, and S. Migliorini, 2023: The effective use of anchor observations in variational bias correction in the presence of model bias. *Quarterly Journal of the Royal Meteorological Society*, **149 (754)**, 1789–1809, https://doi.org/https://doi.org/10.1002/qj.4482, URL https://rmets.onlinelibrary.wiley.com/doi/abs/10.1002/qj.4482, https://rmets.onlinelibrary.wiley.com/doi/pdf/10.1002/qj.4482.

Gaspari, G., and S. E. Cohn, 1999: Construction of correlation functions in two and three dimensions. *Quarterly Journal of the Royal Meteorological Society*, **125 (554)**, 723–757, https://doi.org/10.1002/qj.49712555417.

Gelaro, R., and Coauthors, 2017: The modern-era retrospective analysis for research and applications, version 2 (MERRA-2). *Journal of Climate*, **30 (14)**, 5419–5454, https://doi.org/10.1175/JCLI-D-16-0758.1.

Gibbons, S., T. Kværna, and P. Näsholm, 2019: Characterization of the infrasonic wavefield from repeating seismo-acoustic events. *Infrasound Monitoring for Atmospheric Studies*, Springer, 387–407, https://doi.org/10.1007/978-3-319-75140-5_10.

Gibbons, S. J., and Coauthors, 2015: The European Arctic: A laboratory for seismoacoustic studies. *Seismological Research Letters*, **86 (3)**, 917–928, https://doi.org/10.1785/0220140230.

Green, D. N., and A. Nippress, 2019: Infrasound signal duration: The effects of propagation distance and waveguide structure. *Geophysical Journal International*, **216 (3)**, 1974–1988, https://doi.org/10.1093/gji/ggy530.

Hamill, T. M., J. S. Whitaker, and C. Snyder, 2001: Distance-dependent filtering of background error covariance estimates in an ensemble Kalman filter. *Monthly Weather Review*, **129 (11)**, 2776–2790, https://doi.org/10.1175/1520-0493(2001)129⟨2776:DDFOBE⟩2.0.CO;2.

Hersbach, H., and Coauthors, 2019: Global reanalysis: goodbye ERA-Interim, hello ERA5. *ECMWF Newsletter*, **159**, 17–24, https://doi.org/10.21957/vf291hehd7.

Hintze, J. L., and R. D. Nelson, 1998: Violin plots: A box plot-density trace synergism. *The American Statistician*, **52 (2)**, 181–184, https://doi.org/10.1080/00031305.1998.10480559, URL https://www.tandfonline.com/doi/abs/10.1080/00031305.1998.10480559, https://www.tandfonline.com/doi/pdf/10.1080/00031305.1998.10480559.

Houtekamer, P. L., and F. Zhang, 2016: Review of the Ensemble Kalman Filter for atmospheric data assimilation. *Monthly Weather Review*, **144 (12)**, 4489–4532, https://doi.org/10.1175/MWR-D-15-0440.1.

Hunt, B., E. Kostelich, and I. Szunyogh, 2007: Efficient data assimilation for spatiotemporal chaos: A local ensemble transform Kalman filter. *Physica D: Nonlinear Phenomena*, **230 (1–2)**, 112–126, https://doi.org/10.1016/j.physd.2006.11.008.

Jazwinski, A., 1970: *Stochastic processes and filtering theory*. No. 64, Mathematics in science and engineering, Acad. Press, New York, NY.

Johnson, J. B., J. Anderson, O. Marcillo, and S. Arrowsmith, 2012: Probing local wind and temperature structure using infrasound from volcan villarrica (chile). *Journal of Geophysical Research: Atmospheres*, **117 (D17)**, https://doi.org/10.1029/2012JD017694.

Kalman, R. E., 1960: A new approach to linear filtering and prediction problems. *Transactions of the ASME–Journal of Basic Engineering*, **82 (Series D)**, 35–45, https://doi.org/10.1115/1.3662552.

Kalnay, E., 2003: *Atmospheric Modeling, Data Assimilation and Predictability*. Cambridge University Press, https://doi.org/10.1017/CBO9780511802270.

Khatri, C. G., and C. R. Rao, 1968: Solutions to some functional equations and their applications to characterization of probability distributions. *Sankhyā: The Indian Journal of Statistics, Series A (1961-2002)*, **30 (2)**, 167–180.

Koster, R. D., and Coauthors, 2016: MERRA-2 input observations: Summary and assessment. Tech. rep., National Aeronautics and Space Administration – NASA, Greenbelt.

Kotsuki, S., K. Kurosawa, and T. Miyoshi, 2019: On the properties of ensemble forecast sensitivity to observations. *Quarterly Journal of the Royal Meteorological Society*, **145 (722)**, 1897–1914, https://doi.org/10.1002/qj.3534.

Lalande, J.-M., O. Sèbe, M. Landès, P. Blanc-Benon, R. S. Matoza, A. Le Pichon, and E. Blanc, 2012: Infrasound data inversion for atmospheric sounding. *Geophysical Journal International*, **190 (1)**, 687–701, https://doi.org/10.1111/j.1365-246X.2012.05518.x.

Le Pichon, A., E. Blanc, D. Drob, S. Lambotte, J. Dessa, M. Lardy, P. Bani, and S. Vergniolle, 2005: Infrasound monitoring of volcanoes to probe high-altitude winds. *Journal of Geophysical Research: Atmospheres*, **110 (D13)**, https://doi.org/10.1029/2004JD005587.

Le Pichon, A., E. Blanc, and A. Hauchecorne, 2010: *Infrasound monitoring for atmospheric studies*. Springer Science & Business Media, https://doi.org/10.1007/978-1-4020-9508-5.

Le Pichon, A., E. Blanc, and A. Hauchecorne, 2019: *Infrasound monitoring for atmospheric studies: Challenges in middle atmosphere dynamics and societal benefits*. Springer, https://doi.org/10.1007/978-3-319-75140-5.

Lee, J. C. K., J. Amezcua, and R. N. Bannister, 2022: Hybrid ensemble-variational data assimilation in abc-da within a tropical framework. *Geoscientific Model Development*, **15 (15)**, 6197–6219, https://doi.org/10.5194/gmd-15-6197-2022.

Lei, L., and J. L. Anderson, 2014: Comparisons of empirical localization techniques for serial ensemble Kalman filters in a simple atmospheric general circulation model. *Monthly Weather Review*, **142 (2)**, 739–754, https://doi.org/10.1175/MWR-D-13-00152.1.

Lei, L., J. L. Anderson, and J. S. Whitaker, 2016: Localizing the impact of satellite radiance observations using a global group ensemble filter. *Journal of Advances in Modeling Earth Systems*, **8 (2)**, 719–734, https://doi.org/10.1002/2016MS000627.

Livezey, R., and W. Chen, 1983: Statistical field significance and its determination by monte carlo techniques. 46–59 pp.

43

Lynch, P., 2008: The origins of computer weather prediction and climate modeling. *Journal of Computational Physics*, **227 (7)**, 3431–3444, https://doi.org/10.1016/j.jcp.2007.02.034.

Marty, J., 2019: The IMS infrasound network: current status and technological developments. *Infrasound Monitoring for Atmospheric Studies*, Springer, 3–62.

Nerger, L., T. Janjić, J. Schröter, and W. Hiller, 2012: A unification of ensemble square root Kalman filters. *Monthly Weather Review*, **140 (7)**, 2335–2345, https://doi.org/10.1175/MWR-D-11-00102.1.

Ostashev, V. E., and D. Norris, 2022: Ray tracing in a stratified moving atmosphere: Azimuthal deviation in sound propagation. *The Journal of the Acoustical Society of America*, **151 (1)**, 451–456, https://doi.org/10.1121/10.0009311.

Ott, E., and Coauthors, 2004: A local ensemble Kalman filter for atmospheric data assimilation. *Tellus A*, **56 (5)**, 415–428, https://doi.org/10.1111/j.1600-0870.2004.00076.x.

Palmer, T., 2019: The ECMWF ensemble prediction system: Looking back (more than) 25 years and projecting forward 25 years. *Quarterly Journal of the Royal Meteorological Society*, **145**, 12–24, https://doi.org/10.1002/qj.3383.

Park, J., B. W. Stump, C. Hayward, S. J. Arrowsmith, I.-Y. Che, and D. P. Drob, 2016: Detection of regional infrasound signals using array data: Testing, tuning, and physical interpretation. *The Journal of the Acoustical Society of America*, **140 (1)**, 239–259, https://doi.org/10.1121/1.4954759.

Rennie, M. P., L. Isaksen, F. Weiler, J. de Kloe, T. Kanitz, and O. Reitebuch, 2021: The impact of Aeolus wind retrievals on ECMWF global weather forecasts. *Quarterly Journal of the Royal Meteorological Society*, **147 (740)**, 3555–3586, https://doi.org/10.1002/qj.4142.

Sakov, P., and P. A. Sandery, 2015: Comparison of EnOI and EnKF regional ocean reanalysis systems. *Ocean Modelling*, **89**, 45–60, https://doi.org/j.ocemod.2015.02.003.

Shapiro, S. S., and M. B. Wilk, 1965: An analysis of variance test for normality (complete samples)†. *Biometrika*, **52 (3-4)**, 591–611, https://doi.org/10.1093/biomet/52.3-4.591,

URL https://doi.org/10.1093/biomet/52.3-4.591, https://academic.oup.com/biomet/article-pdf/52/3-4/591/962907/52-3-4-591.pdf.

Shepherd, T. G., I. Polichtchouk, R. J. Hogan, and A. J. Simmons, 2018: Report on stratosphere task force. *ECMWF Tech. Memo*, 1–32, https://doi.org/10.21957/0vkp0t1xx.

Shlyaeva, A., J. S. Whitaker, and C. Snyder, 2019: Model-space localization in serial ensemble filters. *Journal of Advances in Modeling Earth Systems*, **11 (6)**, 1627–1636, https://doi.org/10.1029/2018MS001514.

Simmons, A., and Coauthors, 2020: Global stratospheric temperature bias and other stratospheric aspects of ERA5 and ERA5.1. *ECMWF Tech. Memo*, https://doi.org/10.21957/rcxqfmg0.

Smets, P., J. Assink, A. Le Pichon, and L. Evers, 2016: ECMWF SSW forecast evaluation using infrasound. *Journal of Geophysical Research: Atmospheres*, **121 (9)**, 4637–4650, https://doi.org/10.1002/2015JD024251.

Stober, G., and Coauthors, 2021: Atmospheric tomography using the nordic meteor radar cluster and chilean observation network de meteor radars: network details and 3d-var retrieval. *Atmospheric Measurement Techniques*, **14 (10)**, 6509–6532, https://doi.org/10.5194/amt-14-6509-2021.

Tippett, M. K., J. L. Anderson, C. H. Bishop, T. M. Hamill, and J. S. Whitaker, 2003: Ensemble square root filters. *Monthly Weather Review*, **131 (7)**, 1485–1490, https://doi.org/10.1175/1520-0493(2003)131⟨1485:ESRF⟩2.0.CO;2.

van Leeuwen, P. J., 2009: Particle filtering in geophysical systems. *Monthly Weather Review*, **137 (12)**, 4089 – 4114, https://doi.org/10.1175/2009MWR2835.1, URL https://journals.ametsoc.org/view/journals/mwre/137/12/2009mwr2835.1.xml.

Vera Rodriguez, I., S. P. Näsholm, and A. Le Pichon, 2020: Atmospheric wind and temperature profiles inversion using infrasound: An ensemble model context. *The Journal of the Acoustical Society of America*, **148 (5)**, 2923–2934, https://doi.org/10.1121/10.0002482.

Vitart, F., and A. W. Robertson, 2018: The sub-seasonal to seasonal prediction project (S2S) and the prediction of extreme events. *npj Climate and Atmospheric Science*, **1 (1)**, 3, https://doi.org/10.1038/s41612-018-0013-0.

Vorobeva, E., J. Assink, P. J. Espy, T. Renkwitz, I. Chunchuzov, and S. P. Näsholm, 2023: Probing gravity waves in the middle atmosphere using infrasound from explosions. *Journal of Geophysical Research: Atmospheres*, **128 (13)**, e2023JD038 725, https://doi.org/10.1029/2023JD038725.

Wang, X., C. H. Bishop, and S. J. Julier, 2004: Which is better, an ensemble of positive–negative pairs or a centered spherical simplex ensemble? *Monthly Weather Review*, **132 (7)**, 1590 – 1605, https://doi.org/10.1175/1520-0493(2004)132⟨1590:WIBAEO⟩2.0.CO;2.

Williamson, D., 2007: The evolution of dynamical cores for global atmospheric models. *Journal of the Meteorological Society of Japan. Ser. II*, **85B**, 241–269, https://doi.org/10.2151/jmsj.85B.241.