University of Reading

Listening to the atmosphere: using infrasound observations to infer atmospheric conditions

Article

Published Version

Creative Commons: Attribution 4.0 (CC-BY)

Open Access

Amezcua, J. ORCID: https://orcid.org/0000-0002-4952-8354, Averbuch, G. ORCID: https://orcid.org/0000-0002-0403-9354, Näsholm, S. P. ORCID: https://orcid.org/0000-0001-9107-4002 and Arrowsmith, S. ORCID: https://orcid.org/0000-0002-9150-0363 (2026) Listening to the atmosphere: using infrasound observations to infer atmospheric conditions. Journal of Geophysical Research: Atmospheres, 131 (5). e2025JD044841. ISSN 2169-8996 doi: 10.1029/2025jd044841 Available at https://centaur.reading.ac.uk/129034/

It is advisable to refer to the publisher's version if you intend to cite from the work. See Guidance on citing.

To link to this article DOI: http://dx.doi.org/10.1029/2025jd044841

Publisher: American Geophysical Union (AGU)

All outputs in CentAUR are protected by Intellectual Property Rights law, including copyright law. Copyright and IPR is retained by the creators or other copyright holders. Terms and conditions for use of this material are defined in the End User Agreement.

University of Reading

www.reading.ac.uk/centaur

CentAUR

Central Archive at the University of Reading

Reading's research outputs online

Check for updates

JGR Atmospheres

RESEARCH ARTICLE
10.1029/2025JD044841

Listening to the Atmosphere: Using Infrasound Observations to Infer Atmospheric Conditions

Javier Amezcua[1], Gil Averbuch[2], Sven Peter Näsholm[3,4], and Stephen Arrowsmith[5]

[1]Department of Meteorology, University of Reading, Reading, UK, [2]Woods Hole Oceanographic Institution, Woods Hole, MA, USA, [3]NORSAR, Kjeller, Norway, [4]Department of Informatics, University of Oslo, Oslo, Norway, [5]Southern Methodist University, Dallas, TX, USA

Key Points:

- There are sensor arrays around the world constantly measuring infrasound waves, providing unique long-term data sets of opportunity
- We use indirect observations to update middle atmospheric variables (temperature and winds) which are otherwise sparsely observed
- This is the first work performing data assimilation with a full ray-tracing routine as observation operator and applied to recorded data

Correspondence to:

J. Amezcua,
j.amezcua@reading.ac.uk

Citation:

Amezcua, J., Averbuch, G., Näsholm, S. P., & Arrowsmith, S. (2026). Listening to the atmosphere: Using infrasound observations to infer atmospheric conditions. *Journal of Geophysical Research: Atmospheres*, *131*, e2025JD044841. https://doi.org/10.1029/2025JD044841

Received 10 JUL 2025
Accepted 22 FEB 2026

Abstract The stratosphere and mesosphere are important regions for the prediction of weather at the Earth's surface for medium- and long-range forecasts. The availability of observations in these layers is lower than that of the troposphere, especially for the dynamics. While seeking new observational sources is important, there are existing infrasound data sets that provide indirect observations of these layers to be exploited. Infrasound waves generated at the Earth's surface travel horizontally and vertically through the atmosphere, and can be detected by sensor arrays at ranges of hundreds or thousands of kilometers. These waves are affected by the atmospheric conditions they encounter during propagation, and the integrated contributions can be observed in the recorded observations. Inverse problem techniques can be readily used to extract information from these integrated observations and provide valuable data related to the atmospheric conditions. We exploit observations from controlled ammunition explosions in Oklahoma, which generate waves traveling to 30–50 km altitude before being refracted back to the surface and detected 256 km from the explosion site. As model background we use the 10-member European Reanalysis product, valid 1 hr before the explosions. We use the Modulated Ensemble Transform Kalman Filter to combine these two sources of information and obtain updated atmospheric profiles. The assimilated observations bring the atmospheric profiles closer to those obtained by solely interpolating the reanalysis product to the time of the explosions. The most benefited altitudes are those close to the refraction heights of the infrasound waves, 35–55 km.

Plain Language Summary Observations of stratospheric and mesospheric variables are rare, particularly of winds. There are, however, indirect and underutilized observations which can help constrain variables in these levels. Detonations at the Earth's surface generate infrasound waves which travel horizontally and vertically through the atmosphere. As they travel, they are affected by winds and temperature. The integrated effect is contained in observations taken when these waves are detected at specialized sensor stations. In this work, we use data assimilation techniques to extract information of the atmospheric variables from integrated observations. We take advantage of a unique data set of infrasound recorded from a set of controlled explosions in Oklahoma.

1. Introduction

1.1. The Need to Fill the Upper Atmosphere Observation Gap

The weather we experience on Earth's surface is influenced by what happens in the levels aloft. Processes in both the stratosphere (∼15−50) km and mesosphere (∼50−80) km, and their connection to the troposphere, are important for numerical weather prediction and climate prediction (Charlton-Perez et al., 2013; Kidston et al., 2015; Orsolini et al., 2011). For this reason, the top of current weather prediction models have been extended well into the mesosphere (Polavarapu et al., 2005). Hence, it is important to have accurate observations to initialize and constrain these models.

Improvements in stratospheric wind and temperature modeling have the potential to significantly improve subseasonal-to-seasonal (S2S) prediction at time scales of weeks to months (Baldwin et al., 2003; Domeisen et al., 2020a, 2020b; Mitchell et al., 2013; Scaife et al., 2022; Waugh et al., 2017). For instance, an optimal stratospheric wind representation may improve large-scale weather pattern forecasting like the North-Atlantic Oscillation by 5−10% on average (Charlton-Perez et al., 2021).

Unfortunately, the observational coverage of the stratosphere and mesosphere is nowhere near as dense as that of the troposphere. This has consequences. Some of the largest uncertainties in current reanalysis products occur at

© 2026. The Author(s).
This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

21698996, 2026, 5, Downloaded from https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2025JD044841 by <Shibboleth>-member@reading.ac.uk, Wiley Online Library on [22/06/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

these levels, with the representation of mesospheric winds lagging behind other altitude regions (Baker et al., 2014; Korhonen et al., 2019). Duruisseau et al. (2017) compared ERA-Interim reanalysis winds against (rare) in situ measurements (e.g., from balloon radiosondes) at levels higher than 20 hPa, and found discrepancies of more than 5 m/s Le Pichon et al. (2019) compared reanalysis products against measurements from multiple ground-based instruments. For heights of 40−60 km, they found that the standard deviation of the mean zonal wind difference at times could be larger than 20 m/s. It is clear that information that can be used to better constrain such estimations is of paramount value.

In recent years, there has been a gradual increase in the availability of observations for the middle atmosphere. One of the most important examples is the Aeolus satellite. Launched in 2018 by the European Space Agency (ESA), it carried the first Doppler Wind LiDAR (DWL) in space, designed to provide wind information from the surface up to the lower stratosphere, that is, ∼30 km (Rennie et al., 2021). These observations promise to improve forecasts: simulation experiments with the European Centre for Medium Range Weather Forecast (ECMWF) system showed error reductions of up to 2% for horizontal winds and temperature on days 2–3 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.

While the efforts to expand the observational coverage in the mid and upper atmosphere are encouraging, it is also true that there are existing observation networks which are not being exploited in operational Data Assimilation. The current article focuses on the opportunity of using data collected from infrasound station networks around the world (Assink et al., 2019; Marty, 2019). These data sets contain indirect information of tropospheric, stratospheric and mesospheric variables. There is an interest to maintain these observatories, for example, to ensure the compliance of the Comprehensive Nuclear-Test-Ban Treaty (CTBT). From a research perspective, an objective of the EU-funded ARISE (Atmospheric dynamics Research InfraStructure in Europe) project 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). It is timely to further explore the use of these data sources to improve the stratospheric representation in atmospheric models.

1.2. Information Contained in Infrasound Observations

Infrasound waves are low-frequency sound waves in the range 0.02−20 Hz. These waves are generated by a wide range of sources, which include natural and anthropogenic (Bedard & Georges, 2000). Some sources of infrasound are transient and relatively rare (such as explosions, see Gibbons et al. (2019) and the references therein) while others are continuous and ubiquitous (such as infrasound generated by ocean processes, called microbaroms, see De Carlo et al. (2021) and its references). Due to their low frequency, these waves can propagate through the atmosphere for distances of hundreds to thousands of kilometers, while suffering little dissipation. They can eventually be detected by microbarometers located at ground level, which are typically deployed as arrays for signal enhancement and to measure direction-of-arrival. As they travel through the atmosphere, these waves are affected by the variables of the atmosphere they cross, especially the temperature and wind fields. The integrated effects of these atmospheric variables are encoded in the wavefield recorded at the station where the waves arrive, and features can be extracted from the wavefield that are sensitive to specific atmospheric properties.

Let us focus on three observed quantities. The first is the backazimuth deviation. The great-circle azimuth from the receiver array to the infrasound source (termed the backazimuth) indicates the direction of origin of the source. In the presence of cross-wind (wind perpendicular to the propagation path), the wavefront is advected, making the wave seem to come from a different (apparent) direction (Diamond, 1964). The reader is referred to Blixt et al. (2019) and Amezcua et al. (2020) for a detailed geometrical explanation of this effect. The second observation is travel time. In this case the temperature plays a more direct role since the adiabatic sound speed is proportional to the square root of the temperature:

$$c_{\text{adiabatic}} \propto \sqrt{T}. \tag{1}$$

This makes the wave travel faster in warmer regions than in cooler regions. Furthermore, any headwind the wave encounters will slow its progress, while any tailwind will make it quicker. The effective sound speed is defined as:

$$c_{\text{effective}} = c_{\text{adiabatic}} + v, \tag{2}$$

where v is the horizontal wind speed in the direction of the sound propagation.

Finally, the trace or apparent velocity indicates the elevation angle of arrival of a wavefront. An infinite trace velocity means the wave is arriving completely vertically, while a trace velocity equal to the actual effective sound speed means the wave arrives completely horizontally.

1.3. The Forward Problem, the Inverse Problem, and Our Efforts to Solve the Latter

One can take a given atmospheric profile and predict a set of corresponding observations. This is the so-called *forward* problem. There is a hierarchy of modeling methods to solve this problem that make various levels of approximations to the fundamental fluid dynamic equations describing waves in the atmosphere. These range from simple geometric models (e.g., Blixt et al., 2019), to the full solution of relevant fluid dynamic equations (e.g., Brissaud et al., 2016). In practice, the method suitable for a given problem depends on the propagation physics affecting the observations. A widely used maxim is to use the simplest model that is adequate for a given problem. In this study, we focus on refracted infrasound waves, which can be sufficiently described by 3D ray theory. In particular we use InfraGA (P. S. Blom et al., 2015; P. Blom & Waxler, 2017), a numerical solver to the Eikonal equation developed by Los Alamos National Laboratory.

The so-called inverse problem can be stated as: given a set of observations, which atmospheric profile produced them? Unlike the forward problem, this problem does not have a unique solution, since many different profiles can produce the same observations. There are ways to regularize inverse problems, in particular using so-called data assimilation (e.g., Asch et al., 2016; Stuart, 2010).

In the last few years, a handful of studies have started to explore data assimilation of infrasound observations coming from controlled detonations. While such signals are relatively rare in comparison to the ubiquitous microbaroms (and other continuous or quasi-continuous signals relevant for boundary layer studies), they provide an excellent proving ground for assessing the application of data assimilation of real infrasound data for the middle atmosphere. There are encouraging machine-learning based results pointing in the direction of a future global microbarom-based atmospheric model data assimilation (Vorobeva et al., 2024). In addition, there are studies that address some of the interconnected building blocks needed for a prospective microbarom assimilation framework: source modeling (De Carlo et al., 2021), efficient global infrasonic wave propagation (Brissaud et al., 2022; Cameijo et al., 2025), customized array signal processing (Vorobeva et al., 2021), and atmospheric model benchmarking (Letournel et al., 2024). On the other hand, munition detonations are comparably easy to study since one can consider the source to be a point, and the location and time of the detonation are usually well-known. Amezcua et al. (2020) and Amezcua and Barton (2021) reported the first infrasound-based data assimilation implementations. These used real observations from detonations in Hukkakero (Finland), detected by the ARCES array in Norway. As observation operator they used the geometric model of Blixt et al. (2019), which assumes that the ray trajectory is a perfect isosceles triangle with a prescribed reflection height. Amezcua et al. (2024) improved on this work by using the InfraGA ray-tracing-based wave propagation as an integrated part of the observation operator. This was demonstrated on synthetic data, performing an Observing-Systems Simulation Experiment (OSSE) (Arnold & Dey, 1986). This was based on a set of synthetically generated (yet realistic) atmospheric profiles, and performed experiments with simulated observations.

The purpose of the current work is to perform data assimilation experiments using a full ray-tracing routine and real-world observations of infrasound that has refracted in the stratosphere. In particular, we use observations generated by detonations of old ammunition by the McAlester Army Ammunition Plant (McAAP), Oklahoma (G. Averbuch, Sabatini, & Arrowsmith, 2022). Almost daily, starting just after 16 UTC, the 52 pits are filled with old ammunition and detonated at intervals of 20 s. A sequence consists of two test detonations, a one-minute break, and the remaining detonations. A two-minute break separates the detonation sequences from the two areas. A microbarometer array was deployed approximately 4 km at an azimuth of 35° from the detonation sites. This array acts as a source reference. In order to detect stratospheric signals, an array was deployed at a distance of 256 km from the detonation sites at an azimuth of 269°. The waves take roughly 13 min to travel from source to receiver. We focus on the detonations of the 21 July 2021. Further details on the geographical setting can be found in Figure 1 of G. Averbuch, Sabatini, and Arrowsmith (2022).

21698996, 2026, 5, Downloaded from https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2025JD044841 by <Shibboleth>-member@reading.ac.uk, Wiley Online Library on [22/06/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

21698996, 2026, 5, Downloaded from https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2025JD044841 by <Shibboleth>-member@reading.ac.uk, Wiley Online Library on [22/06/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

The rest of this article is organized as follows. Section 2 provides a description of the methodology. Section 3 describes our experimental setup. In Section 4 we present the results of our assimilation experiments. Section 5 contains the summary and concluding remarks of the work.

2. Methodology: Regularizing the Inverse Problem

Data assimilation (DA) is a mathematical discipline which combines information from different and imperfect sources to determine the value of the so-called state variables in a system of interest. Typically, these sources are a model (also known as background) and observations, both with uncertainties. DA applies ideas from control theory and statistical filtering (Jazwinski, 1970), and uses Bayesian methods to regularize inverse problems (Stuart, 2010). The reader is referred to, for example, Asch et al. (2016) for a comprehensive introduction to this topic. We use a particular formulation of the ensemble Kalman filter (EnKF; Evensen, 2006). Here we introduce its analysis step in the simplest case, then provide a toy example, and finally describe the actual application in our setting.

2.1. Combining Information

Imagine you have two imperfect sources of information for the same state variable x: the background x^b (it can come from a forecast) and the observation y. Both contain an additive error:

$$\begin{aligned} x^b &= x^{\text{true}} + \epsilon, \\ y &= x^{\text{true}} + \eta. \end{aligned} \tag{3}$$

These errors follow a Gaussian distribution. Both are unbiased, meaning that their expected value is zero: $\mathrm{E}[\epsilon] = \mathrm{E}[\eta] = 0$. Their variances are $\mathrm{Var}[\epsilon] = b^2$ and $\mathrm{Var}[\eta] = r^2$. The unbiased-ness of both errors is an essential assumption in classic data assimilation (e.g., Kalnay, 2003). Biased errors can be handled by estimating their mean, in a process called parameter estimation.

One can construct a new estimator (denoted analysis) as a linear combination of the two sources:

$$x^a = c_1 x^b + c_2\, y. \tag{4}$$

Optimal interpolation (OI) is at the heart of the analysis step of the EnKF and other DA methods. To find the coefficients of (Equation 4) which guarantee minimum variance of x^a, OI minimizes the quadratic cost function

$$J(x) = \frac{1}{2}\frac{\left(x - x^b\right)^2}{b^2} + \frac{1}{2}\frac{(y - x)^2}{r^2}. \tag{5}$$

This equation has two terms. The first one indicates the closeness to the background, weighted by its uncertainty, while the second term indicates the closeness to the actual observation, weighted by its uncertainty. This simple minimization problem yields:

$$x^a = \frac{\frac{1}{b^2}}{\frac{1}{b^2} + \frac{1}{r^2}} x^b + \frac{\frac{1}{r^2}}{\frac{1}{b^2} + \frac{1}{r^2}} y. \tag{6}$$

This expression is readily interpretable. The factor $1/b^2$ is the precision of the background and $1/r^2$ the precision of the observation. Hence, each coefficient in the linear combination is the relative precision of x^b and y, correspondingly. What happens when many state variables combine to produce a single observation? This is explained next.

2.2. Integrated Observations

Here we provide a toy example with only one observation, but with several state variables. This setting keeps important aspects of the full ray-tracing problem we handle later. In particular, it yields integrated (or summed) observations, making its inverse problem ill-posed.

21698996, 2026, 5, Downloaded from https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2025JD044841 by <Shibboleth>-member@reading.ac.uk, Wiley Online Library on [22/06/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Consider a vertical column with three levels and one variable per level. These variables are organized in a column vector $\mathbf{x}$, and the dimension of the state space is $N_x = 3$. We would like to know the values of these variables, but we cannot observe each value individually, only a weighted sum of them. The so-called observation equation of our problem becomes:

$$y = \mathbf{H}\mathbf{x} + \eta, \tag{7}$$

where $\mathbf{H}$ is known as observation operator, which in this case is a row vector of weights:

$$\mathbf{H} = [w_1, w_2, w_3]. \tag{8}$$

The *forward* problem consists of having a given value of $\mathbf{x}$ and obtaining the corresponding value of y. An illustration of this process is presented in Figure 1, where we let the weights $w_i = 1$ for simplicity. Four different columns are presented in the top row, with the corresponding observations in the bottom row. Information is lost when the observation is taken. Columns 1 and 2 are different: $\mathbf{x} = [0,0,0]^\top$ and $\mathbf{x} = [-1,0,1]^\top$, yet they yield the same observation $y = 0$. Columns 3 and 4 are different: $\mathbf{x} = [0,0,1]^\top$ and $\mathbf{x} = [-1,1,1]^\top$, yet they yield the same observation $y = 1$.

The *inverse* problem is as follows: given an observation y, determine the column $\mathbf{x}$ that generated it. We just showed that multiple columns can generate the same observation, so this inverse problem is ill-posed, that is, it does no have a unique solution. To have a unique solution, this problem needs to be *regularized*, which in practical terms means to provide a background $\mathbf{x}^b$ to compare against, and provide uncertainties for the observation and the background. The question becomes: given an observation y, determine the column $\mathbf{x}$ that generated it, *and* is closest to the background column $\mathbf{x}^b$.

In this multidimensional case, the solution of this problem is given by the vector $\mathbf{x}^a$ which minimizes the quadratic cost function:

$$J(\mathbf{x}) = \frac{1}{2}\left(\mathbf{x} - \mathbf{x}^b\right)^\top \mathbf{B}^{-1}\left(\mathbf{x} - \mathbf{x}^b\right) + \frac{1}{2}\frac{(y - \mathbf{H}\mathbf{x})^2}{r^2}. \tag{9}$$

Equation 9 is the multi-dimensional equivalent to (Equation 5). In this, $\mathbf{B}$ is the background error covariance matrix. This symmetric matrix contains the background error variances in its diagonal, and the off-diagonal elements contain covariances between variables, namely:

$$\mathbf{B} = \begin{bmatrix} b_{1,1} & b_{1,2} & b_{1,3} \\ b_{2,1} & b_{2,2} & b_{2,3} \\ b_{3,1} & b_{3,2} & b_{3,3} \end{bmatrix}. \tag{10}$$

For a linear observation operator, there is a simple analytic solution to this minimization problem. This is known as the analysis equation of the Kalman Filter (KF) (Kalman, 1960; Kalman & Bucy, 1961). Namely,

$$\mathbf{x}^a = \mathbf{x}^b + \frac{\mathbf{B}\mathbf{H}^\top}{\gamma^2}\Delta y, \tag{11}$$

where

$$\Delta y = y - \mathbf{H}\mathbf{x}^b \tag{12}$$

is the observational increment, that is, the difference between the actual observation and that predicted by the background, and γ^2 is known as the total covariance of the observations given by

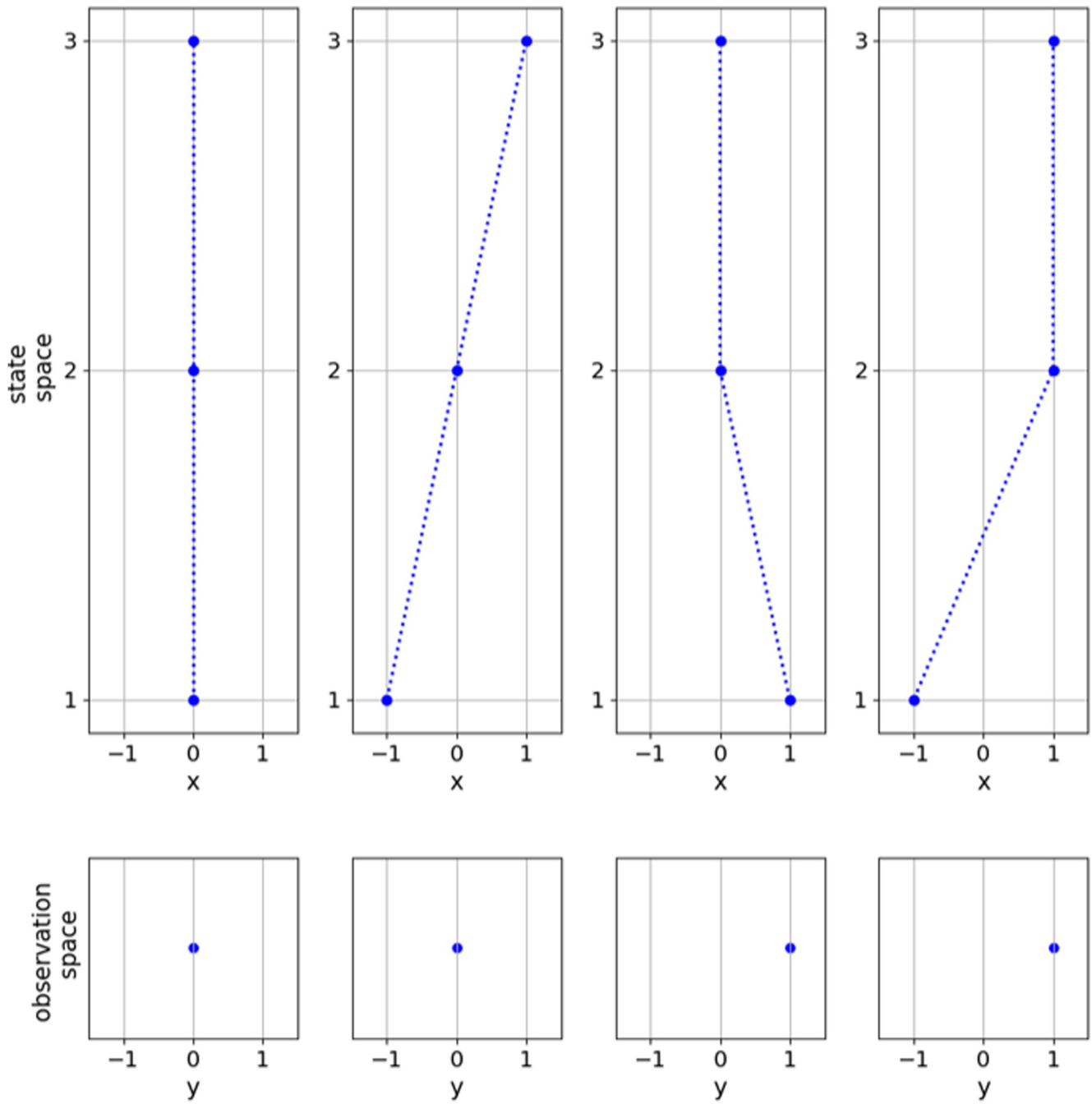

Figure 1. Illustration of the forward problem in a simple toy model. The observation (bottom row) is simply the sum of the three grid points in each column (top row). For this example, the three weights are equal to 1. Different columns can result in the same observation.

$$\gamma^2 = \mathbf{HBH}^{\top} + r^2 = \sum_{i=1}^{3}\sum_{i=1}^{3} w_i w_j b_{i,j} + r^2. \tag{13}$$

Equation 11 is the multidimensional equivalent to (Equation 6), and it also has a very intuitive form. The analysis value is simply the background value (1st term), plus a correction coming from the observational difference (2nd term). The matrix coefficient of the second term is known as the Kalman gain, and it has the role of distributing the

21698996, 2026, 5, Downloaded from https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2025JD044841 by <Shibboleth>-member@reading.ac.uk, Wiley Online Library on [22/06/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

observational increment to the state variables. The way this matrix operates becomes clear if we write (Equation 11) explicitly for our example

$$\begin{bmatrix} x_1^a \\ x_2^a \\ x_3^a \end{bmatrix} = \begin{bmatrix} x_1^b \\ x_2^b \\ x_3^b \end{bmatrix} + \begin{bmatrix} w_1 b_{1,1} + w_2 b_{1,2} + w_3 b_{1,3} \\ w_1 b_{2,1} + w_2 b_{2,2} + w_3 b_{2,3} \\ w_1 b_{3,1} + w_2 b_{3,2} + w_3 b_{3,3} \end{bmatrix} \frac{\Delta y}{\gamma^2}. \tag{14}$$

From Equation 14 we can note that even if the weight of a variable $w_i = 0$ is null in the sum, it can still be affected by Δy if it has a nonzero covariance b_{ij} with a variable whose weight $w_i \neq 0$. This means that unobserved variables can be informed by observed ones.

Figure 2 illustrates the solution to the inverse problem in our toy system with weights equal to 1. In this case we have three rows: state space (top), observation space (middle) and background covariance matrix (bottom). In both cases we have a background $\mathbf{x}^b = [0,0,0]^\top$ (blue lines), yielding a predicted observation $\mathbf{H}\mathbf{x}^b = 0$. In both cases we assimilate a given observation $y = 2$. In the left column $\mathbf{B}$ is a diagonal matrix with the same variance for each state variable, $b_{i,i} = 1$. Hence, the increment in each variable is the same. In the right column $\mathbf{B}$ is still diagonal, but the variances are: $b_{1,1} = 3$, $b_{2,2} = 2$, $b_{3,3} = 1$. For this reason, the changes in the variables are not the same. The variable with largest background uncertainty (x_1) is the one whose value changes the most, while the opposite happens for x_3.

2.3. Ensemble Kalman Filters

So far, we have worked under two assumptions in our examples. (a) We considered that we have access to an estimator $\mathbf{x}^b$ and its covariance matrix $\mathbf{B}$. This is often not the case. (b) We considered the observation operator to be linear. In realistic problems, for example, for satellite observations or the infrasound transmission we study, the observation operators are nonlinear.

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 the observation operator is not too far from linear (e.g., Jazwinski, 1970). It requires the computation and storage of the linearization of the observation operator, which is often cumbersome.

Ensemble Kalman filters (EnKFs; e.g., Houtekamer & Zhang, 2016; Nerger et al., 2012) are sample-based implementations of the EKF. They are fed with a family of M background states (known as ensemble) denoted in the form of a matrix $\mathbf{X}^b$, where each column of the matrix is a single forecast. From $\mathbf{X}^b$, EnKFs compute the sample mean and covariance as estimators of $\mathbf{x}^b$ and $\mathbf{B}$, respectively. They avoid the linearization of the observation operator by mapping each ensemble member to observation space and creating the matrix $\mathbf{Y}^b$, where each column is a set of predicted observations. Then, they readily can compute the necessary sample cross-covariances between $\mathbf{X}^b$ and $\mathbf{Y}^b$ needed.

There are several "flavors" of the EnKF, see, for example, Tippett et al. (2003), Nerger et al. (2012) for a taxonomy. They all share the same objective: to transform the individual members of the background ensemble $\mathbf{X}^b$ into individual members of the analysis ensemble $\mathbf{X}^a$. This is done by implementing the Kalman equations in clever, and sometimes approximate, ways. Symbolically, we have the following mapping between input and output:

$$\{\mathbf{X}^b, \mathbf{Y}^b, \mathbf{y}, \mathbf{R}\} \rightarrow \{\mathbf{X}^a\}, \tag{15}$$

where $\mathbf{y}$ is the vector of actual observations involved in the process, and $\mathbf{R}$ is the observational error covariance matrix. $\mathbf{R}$ is often considered to be prescribed.

Amezcua et al. (2024) applied the symmetric form of the Ensemble Transform Kalman Filter (ETKF; Bishop et al., 2001; Wang et al., 2004) and its localized version in model space, which is known as Modulated ETKF (Bishop et al., 2017; Bishop & Hodyss, 2009). This is the method we apply in this work. The METKF is commonly used in satellite DA. As in our problem, satellite observations are integrated. It is therefore impossible

21698996, 2026, 5, Downloaded from https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2025JD044841 by <Shibboleth>-member@reading.ac.uk, Wiley Online Library on [22/06/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

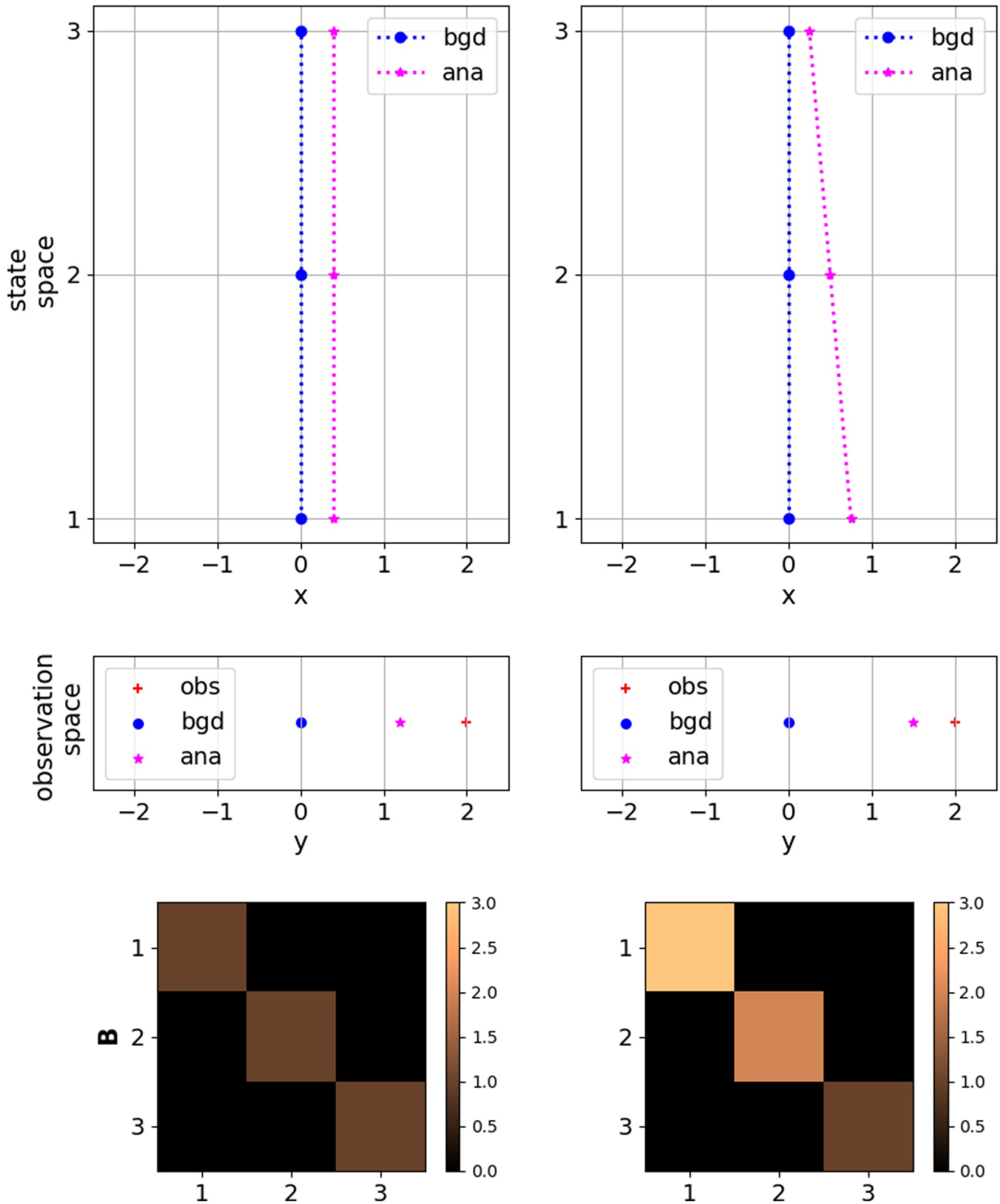

Figure 2. Solution of a simple inverse problem using the analysis step of the Kalman Filter. For both columns the background is zero everywhere, and we assimilate an observation with value $y = 2$. The analysis values depend on the corresponding background error covariance matrix of the problem.

21698996, 2026, 5, Downloaded from https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2025JD044841 by <Shibboleth>-member@reading.ac.uk, Wiley Online Library on [22/06/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

21698996, 2026, 5, Downloaded from https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2025JD044841 by <Shibboleth>-member@reading.ac.uk, Wiley Online Library on [22/06/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

to define a localization distance between state variables and observations (Campbell et al., 2010). The METKF works by artificially extending the ensemble size to obtain an ensemble that prodices (at least approximately) a prescribed (localized) covariance matrix. A detailed description of this algorithm and its application to the infrasound problem can be found in Section 3 of Amezcua et al. (2024).

3. Experimental Setup: Assimilation of Infrasound Observations

We move into the assimilation of infrasound observations. In this case, the state variables are the temperature, zonal and meridional winds over all the N_z vertical levels of an atmospheric profile:

$$\mathbf{x} = [T_1, \ldots, T_{N_z}, u_1, \ldots, u_{N_z}, v_1, \ldots, v_{N_z}]^\top. \tag{16}$$

Other variables such as pressure and density are not considered in this work since they have a negligible impact in the ray-tracing process.

As in previous works (Amezcua et al., 2020, 2024; Blixt et al., 2019), we considered the atmosphere to be range-independent (i.e., temperature and wind fields vary only as a function of height) for the propagation distance of ∼256 km. G. Averbuch, Sabatini, and Arrowsmith (2022) also used this consideration for the modeling after noting that over this given distance, atmospheric specifications did not change much. They also showed that atmospheric changes in time scales of tens of seconds affect the acoustic signals. However, as first step, in this work we will consider it to be static (i.e., no time evolution) during the ∼20 minutes transmission time.

As the background, we use the ensemble ERA5 (European Centre for Medium-Range Weather Forecast Reanalysis v5; Hersbach et al., 2019) model output at 15 UTC, which is the product time closest to the considered detonation times. The top row of Figure 3 displays, from left to right, the temperature, zonal wind, and meridional wind for each of the 10 ERA5 ensemble members (lines with different colors). The temperatures closely agree among the ensemble members, while there is noticeable discrepancy in the winds especially above 40 km. The bottom row has two panels. The left panel shows the computation of the effective sound speed (ESS), which includes effects of both temperature and tail wind. These profiles show an atmospheric duct starting at around 40 km high, suggesting waves will be refracted at this level. Finally, ray tracing is performed with each one of the ensemble members. The resulting ray-paths are illustrated in the middle panel of the bottom row, where indeed the refraction heights are found roughly between 43 and 46 km altitude.

In order to assess the quality of the DA results, we would ideally need independent observations of all the state variables at the assimilation time. Since we do not have this, it is necessary to find a proxy. We interpolate the ERA5 product to 16 UTC, which is the start of the munitions detonation sequence. We label this ensemble "the reference" and we consider its ensemble mean to be the closest to the true profile we are trying to infer. This assumption of course limits our validation to the quality of the ensemble ERA5 product.

The observations are taken from detections at the MCB array (proxy for source) and the WCT array (receiver). All the results are based on the WCT array. The MCB array was in place only to capture the source function and timing. Since it is within several kilometer from the test site, it does not provide any information on the stratosphere. More details of the arrays can be found in Section 1.3, as well as in G. Averbuch, Sabatini, and Arrowsmith (2022).

Figure 4 shows the raw Fisher ratio (an indicator of the signal-to-noise ratio) recorded in both stations starting at 16 UTC. Individual peaks are identified and associated with each one of the 51 detonations that happened on the day. The backazimuth and travel times are measured using array processing at the instants of identified peaks at WCT. The backazimuth deviation is found as the difference of the recorded backazimuth and the geographical backazimuth. For travel times, we pair the peaks at both sites and take the differences in observed times. As in G. Averbuch, Sabatini, and Arrowsmith (2022), 11.4 s are added to the travel time to account for the fact that MCB is located 4 km away from the actual explosion site.

Figure 5 compares the predicted observations with the actual observations. The blue and black dots represent the result of applying ray-tracing to the background ensemble members valid at 15 UTC and 16 UTC respectively. The red dots show the actual 51 observations obtained as described before. The x-axis represents the time at which

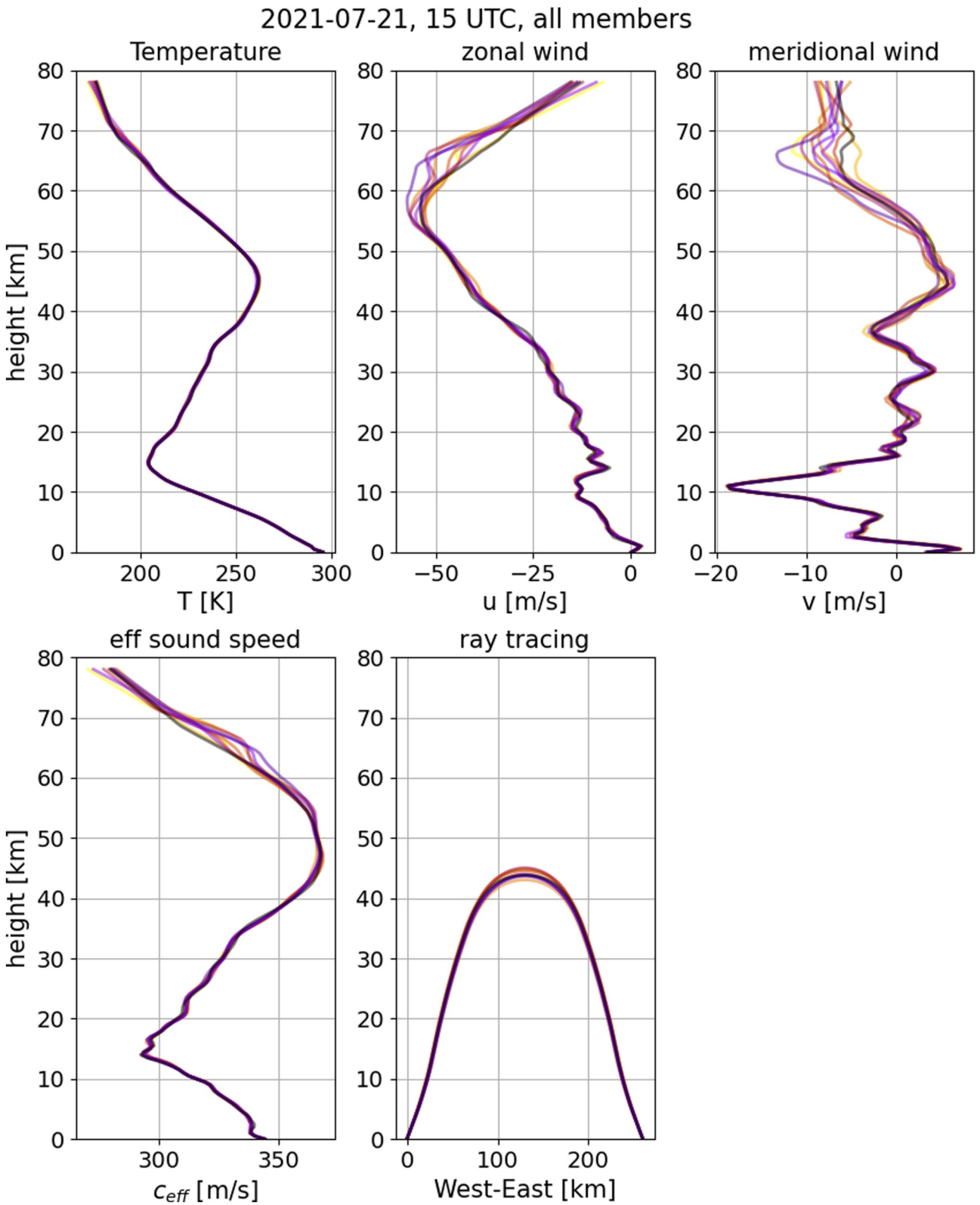

Figure 3. From left to right: atmospheric profiles for temperature, zonal and meridional winds, effective sound speed, and ray tracing. Note that each panel contains 10 different lines.

21698996, 2026, 5, Downloaded from https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2025JD044841 by <Shibboleth>-member@reading.ac.uk, Wiley Online Library on [22/06/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

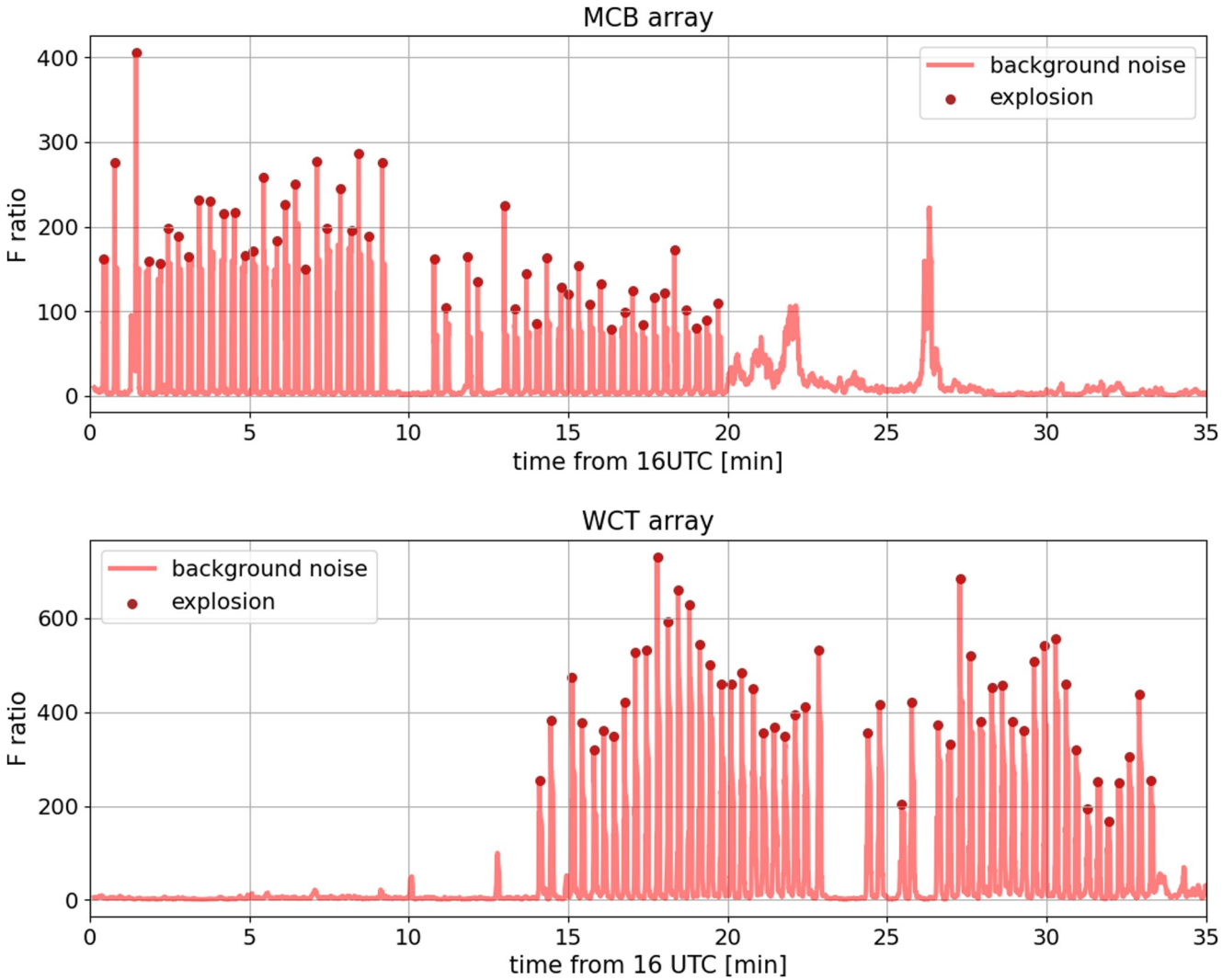

Figure 4. Raw observations at the MCB array (proxy for the source) in the top panel, and at the WCT array (receiver) in the bottom panel, on 21 July 2021. Individual peaks of the Fisher ratio are associated to the 51 detonations of the day (shown as dots).

observations occur, measured in minutes from 16 UTC. Note that the position of the predicted observations in the x-axis is arbitrary.

The backazimuth deviation predicted by the background is close to zero, while the observed deviations can come close to 2°. The predicted travel times are close to 835 s, while the observed times are less (indicating faster wave propagation than predicted), ranging from 822 to 833 s. Finally, the predicted trace velocities are close to 365 m/s, while the observed ones are consistently grouped around 350 m/s. Given the differences between predicted and observed quantities, we expected the DA to have an effect.

4. Results of the DA Experiments

4.1. First Experiments

In our first experiments we use the ETKF with the raw ensemble. Based on the observations, and assuming that the differences in position of each explosion are negligible relative to the distance to WCT, we prescribe the observational error covariance as:

$$\mathbf{R} = \mathrm{diag}\left[(2\ \mathrm{deg})^2, (10\ \mathrm{sec})^2, (5\ \mathrm{m/s})^2\right]. \quad (17)$$

21698996, 2026, 5, Downloaded from https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2025JD044841 by <Shibboleth>-member@reading.ac.uk, Wiley Online Library on [22/06/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

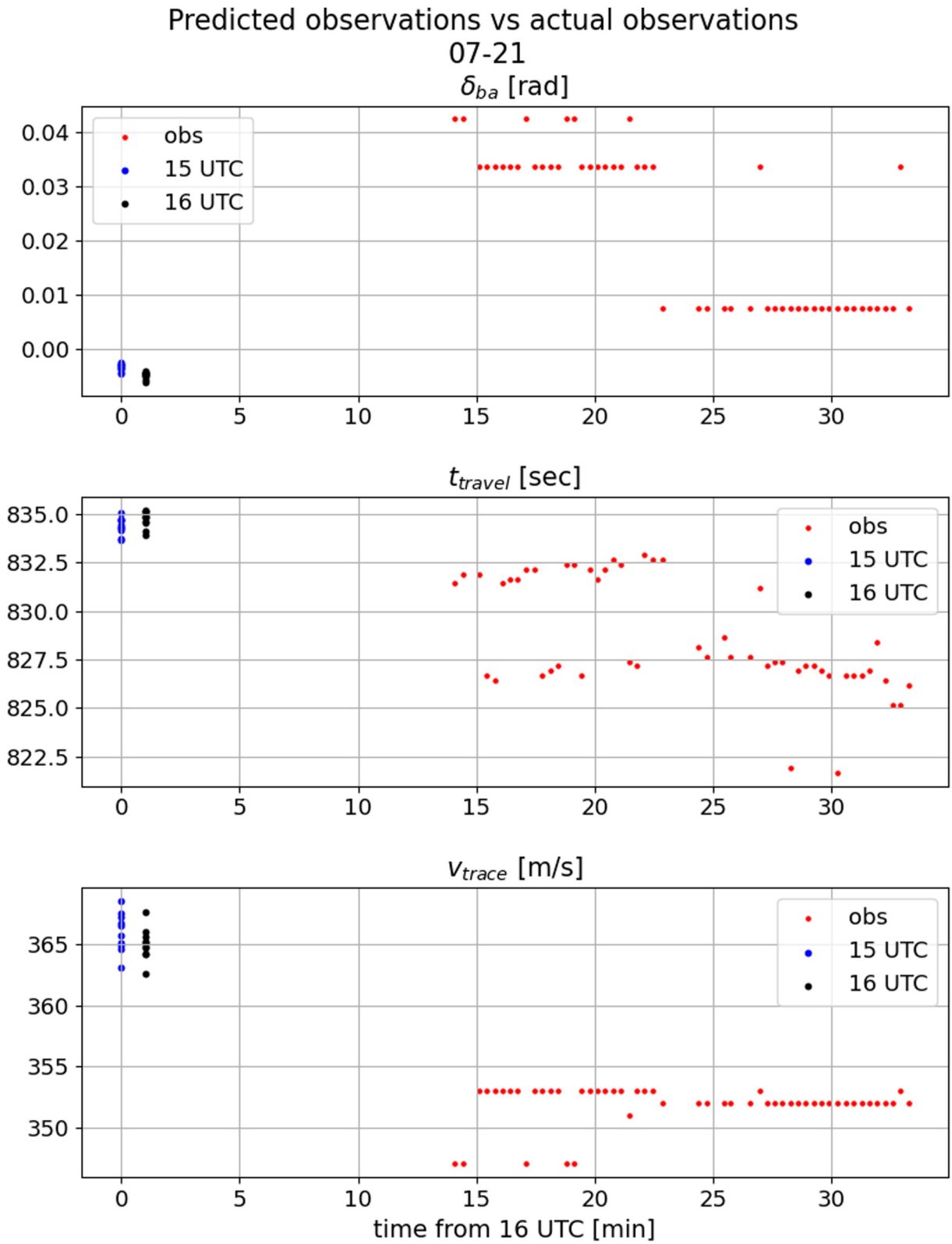

Figure 5. Comparison of predicted observations from the ERA5 ensemble reanalysis (blue and black dots) against the actual observations (red dots). The *x* axis in each panel corresponds to the timing of the observations.

21698996, 2026, 5, Downloaded from https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2025JD044841 by <Shibboleth>-member@reading.ac.uk, Wiley Online Library on [22/06/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

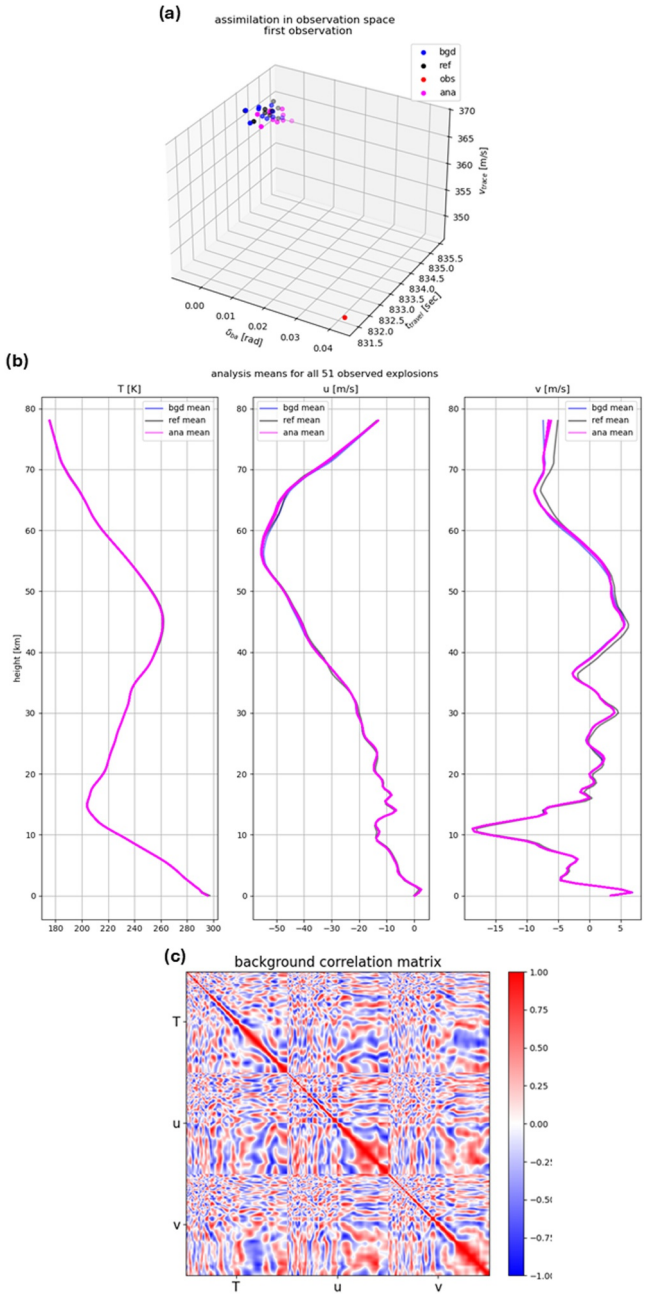

Figure 6. Results of the assimilation experiment using the raw ensemble. Panel (a) shows the results in observation space, panel (b) shows the results in state space, and panel (c) shows the sample covariance coming from the 10-member ensemble.

The results of this experiment are shown in Figure 6. This figure has three panels. Panel (a) shows the quantities in observation space; since this space is three-dimensional it is easy to visualize. The blue dots denote the predicted observations from the background, the red dot is a given observation (the first of the 51) and the magenta dots correspond to the analysis. The reference values are shown in black. In this space, the analysis values are slightly closer to the observation than the background, especially the trace velocity. It would seem that the DA process is fulfilling its purpose.

We then move to the state space (panel b), to see the analysis profiles yielded by the DA. The three subpanels show (from left to right): temperature, zonal and meridional wind. The blue line shows the background ensemble mean, and the black line the reference mean. In magenta, we plot the analysis means coming from assimilating the 51 observations independently. It is very difficult to see any deviation of the analysis from the background. This implies that the observational error we used was too large, and we need to artificially reduce it to increase the effect of the observations.

In panel (c) we show the background correlation matrix generated by the sample (we show correlation and not covariance to focus on the structures and avoid them being masked by the different units of temperatures and winds). This is a 3-by-3 block symmetric matrix with 157 times 157 elements in each block. Not unexpectedly, this is a very noisy estimator–we cannot ask much of a 10-dimensional sample when trying to generate a 147×147 covariance matrix. This means that if we deflate the observational error (as we suggested before), the resulting profiles could result noisy. Therefore we will need to use localization for our next experiments.

4.2. Applying Localization and Inflation

Localization and inflation/deflation are two techniques commonly used in tandem with EnKFs, which allow them to overcome sampling issues with small ensemble sizes (which is our case). Localization is based on the idea that, instantaneously, far away state variables should not have an effect on each other (Anderson, 2001; Hamill et al., 2001). Hence, one can taper a covariance matrix to keep elements coming from close variables, and eliminate the elements coming from faraway variables. The reader is referred to, for example, Anderson (2012) and Farchi and Bocquet (2019) for ample discussions on the topic.

Figure 7 shows the localized versions of the background error correlation matrix under different localization radii (increasing from left to right, from top to bottom). To create the localization matrix, we use the well-known function from Gaspari and Cohn (1999), which is a compact-support approximation to the Gaussian.

The Modulated Ensemble Transform Kalman Filter (METKF) creates an artificial ensemble with a larger size than that of the original ensemble, which by design produces a desired localized covariance matrix. The creation of this ensemble is done via a modulation (Khatri-Rao) product of the original ensemble members, and the eigenvectors coming from a truncated eigen-decomposition of a desired matrix. In our experiments, we used a modulated ensemble of $N_e = 150$ members (meaning we keep 15 eigenvectors) to create matrices with vertical localization radii of $\{3, 5, 7, 9, 11, 13\}$ km. Starting at 5 km, this choice of modulated ensemble size ensures that more than 95% of the total variance in the original localization matrix is kept.

Inflation/deflation involves shrinking or increasing the values in either **B** or **R** to alter the relative importance of the background and observations when computing the analysis. The reader is referred to, for example, Greybush

21698996, 2026, 5, Downloaded from https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2025JD044841 by <Shibboleth>-member@reading.ac.uk, Wiley Online Library on [22/06/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

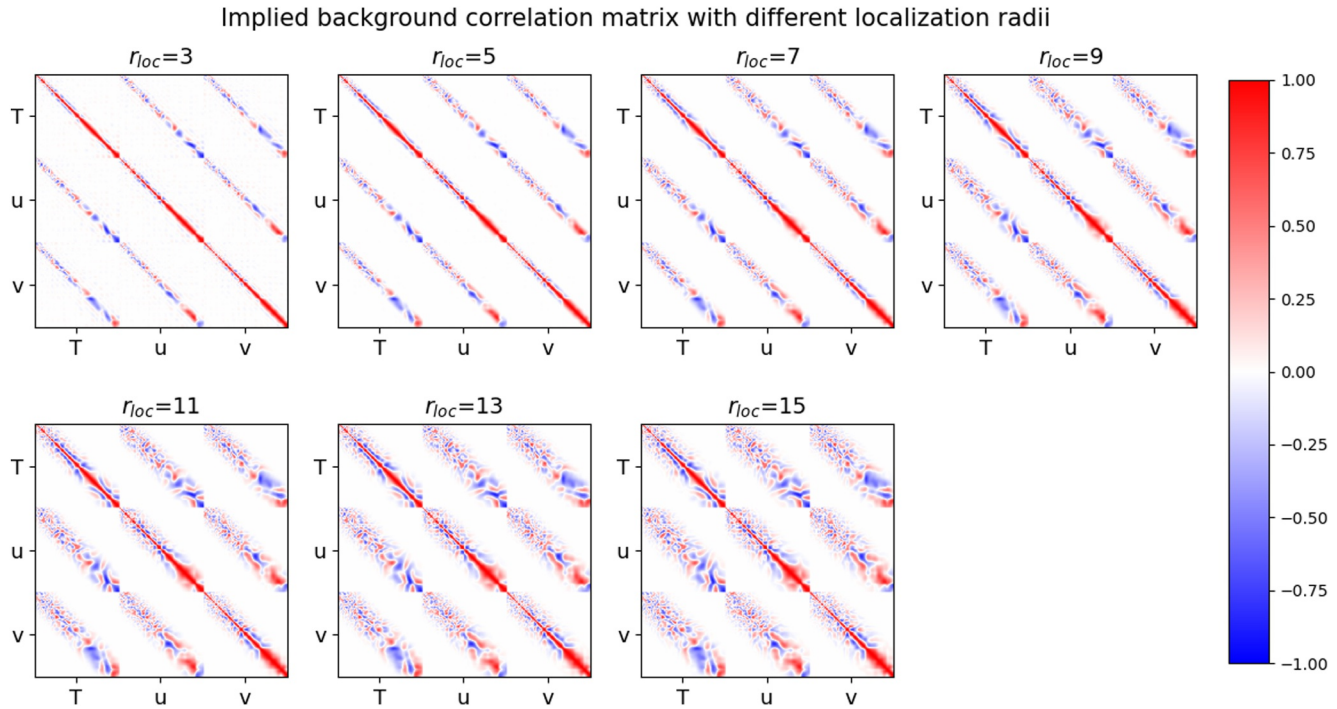

Figure 7. Effects of localization in the background error correlation matrix. We show localization radii of increasing length (left to right).

et al. (2011) for a detailed discussion on this technique. In our experiments, we tune a deflation factor for **R** such that:

$$\mathbf{R} \rightarrow \frac{\mathbf{R}}{\rho}. \tag{18}$$

If $\rho < 1$, the inflation error grows and the observations are ignored more and more by the DA analysis step. When $\rho = 1$ we recover the prescribed covariance matrix. When $\rho > 1$ and as it increases, the error in the observation decreases and hence the DA system gives them more weight for the final analysis values. In our experiments, we start with the value of **R** prescribed in (Equation 17). We then divide this matrix by the following factors $\rho = \{0.5, 1, 1.5, \cdots 10\}$.

Finding optimal values of ρ and λ is the result of tuning. For this process, considering the interpolated 16 UTC values as reference is particularly useful when we perform a grid search for the optimal values. We can search for the combination $\{\rho, \lambda\}$ which minimizes the analysis error with respect to the reference. To assess this error we have two options. On one hand, for each physical variable we can compute the Mean Squared Error (MSE) of the analysis mean with respect to the reference mean as:

$$\mathrm{MSE}^a = \frac{1}{N_{\mathrm{levs}}} \sum_{n=1}^{N_{\mathrm{levs}}} \left(\bar{x}_n^a - \bar{x}_n^{\mathrm{ref}}\right)^2. \tag{19}$$

On the other hand, to combine the effects of the three variables we can use the so-called dry energy norm (DEN; Talagrand, 1981) for the analysis with respect to the reference.

$$\mathrm{DEN}^a = \sum_{n=1}^{N_{\mathrm{levs}}} \left[\frac{1}{2}\left(\bar{u}_n^a - \bar{u}_n^{\mathrm{ref}}\right)^2 + \frac{1}{2}\left(\bar{v}_n^a - \bar{v}_n^{\mathrm{ref}}\right)^2 + \frac{c_p}{2} \frac{\left(\bar{T}_n^a - \bar{T}_n^{\mathrm{ref}}\right)^2}{\bar{T}_n^{\mathrm{ref}}} \right]. \tag{20}$$

We compute this for each one of the 51 observations and take an average over all of them, as well as the minimum and maximum values.

21698996, 2026, 5, Downloaded from https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2025JD044841 by <Shibboleth>-member@reading.ac.uk, Wiley Online Library on [22/06/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

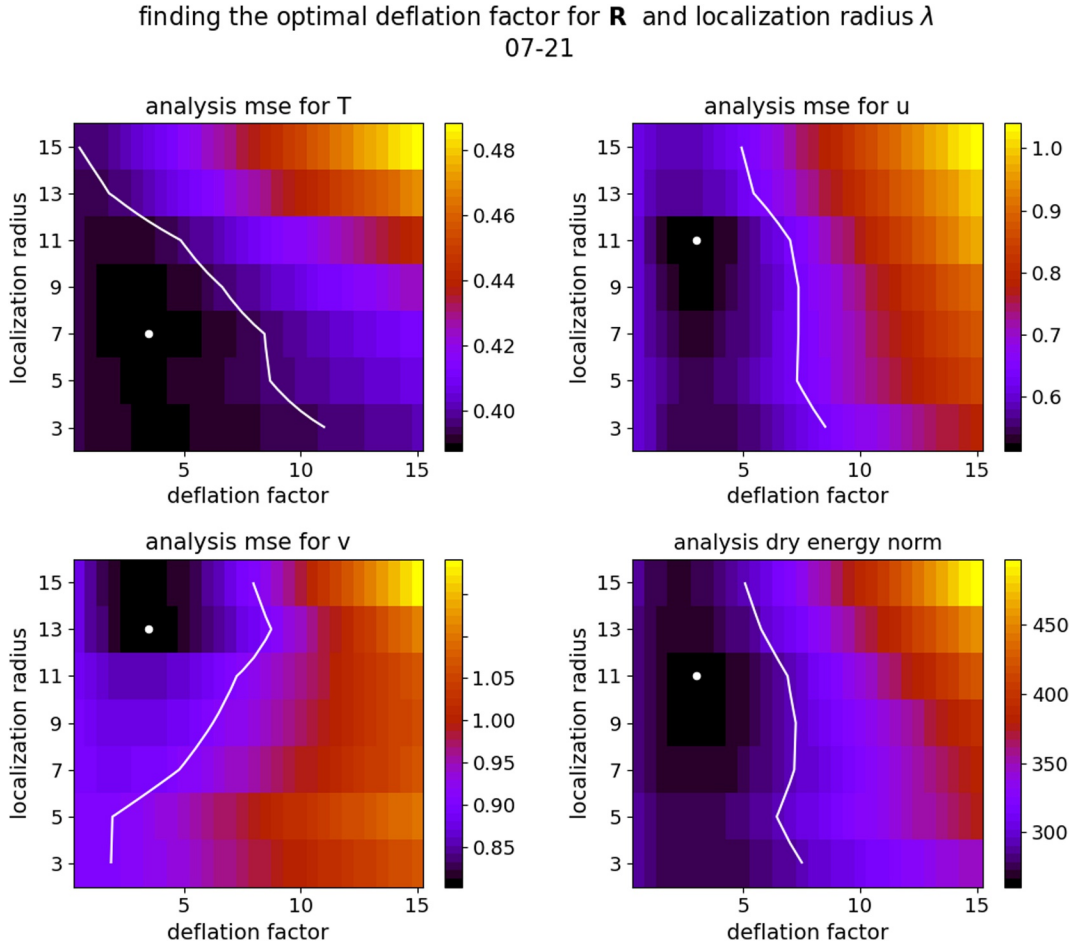

Figure 8. Analysis MSE (for temperatures and wind) and analysis DEN as a function of deflation factor (horizontal axis) and localization radius (vertical axis). In each panel, the white line denotes the background RMSE/DEN. The white dots denote the minimum values.

Figures 8 and 9 show the results of this search process. Figure 8 shows the resulting averaged analysis MSE and DEN (bottom right) from analysis means generated by different combinations of ρ (horizontal axis) and λ (vertical axis). The minimization problems seems to be convex. This means there is a unique global minimum, or at least a unique local minimum within the search area. This renders a unique solution to the minimization problem and avoids ambiguity. We draw the background RMSE/DEN with a white line, everything to the left of this line indicates that the assimilation has been successful. In all cases we are able to find a global minimum, which is indicated with the white dot. In the end we choose DEN, where the minimum corresponds to $\lambda_{\text{opt}} = 11$ km and a deflation factor $\rho_{\text{opt}} = 3$. The latter corresponds to an $\mathbf{R}$:

$$\mathbf{R}_{\text{opt}} = \text{diag}\left[(0.6324\ \text{deg})^2, (3.1622\ \text{sec})^2, (1.5811\ \text{m/s})^2\right]. \tag{21}$$

Figure 9 assesses the variability in this process between different observations. In this figure, we have fixed $\lambda_{\text{opt}} = 11$ and we vary the deflation parameter. The first three panels show analysis MSE for each variable, while the last one (bottom right) shows the analysis DEN. In each panel we have denoted the background MSE/DEN with a horizontal cyan line. For the analysis MSE/DEN, we have three magenta lines. The dotted lines show the maximum and minimum MSE/DEN among all the observations, and the solid line shows the averaged MSE/DEN over all the observations. We see little spread around the minimum average value, and again we see that the minimizer for DEN is $\rho_{\text{opt}} = 3$.

21698996, 2026, 5, Downloaded from https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2025JD044841 by <Shibboleth>-member@reading.ac.uk, Wiley Online Library on [22/06/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

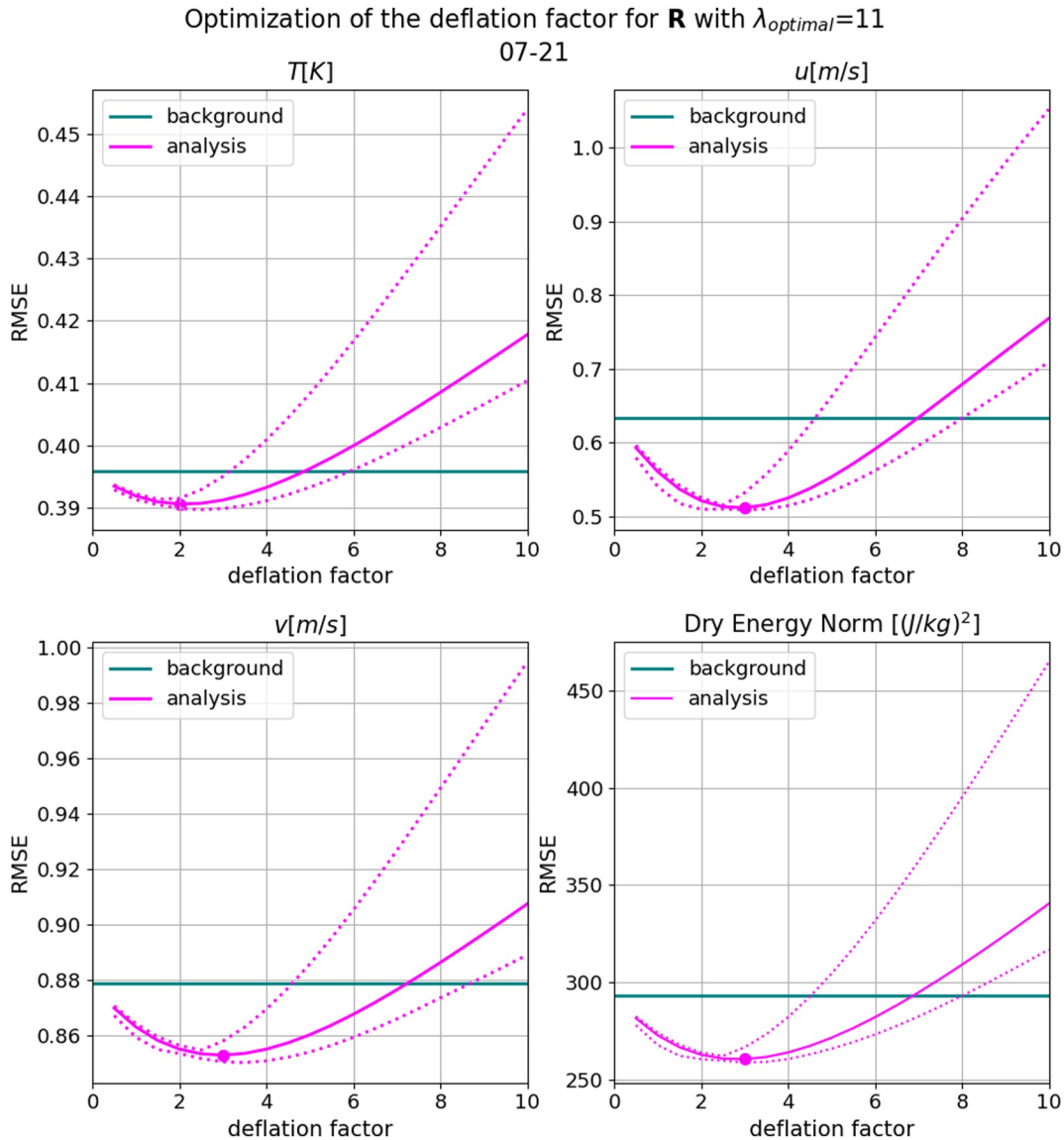

Figure 9. Analysis MSE and DEN as a function of deflation factor (horizontal axis) for a fixed localization radius $\lambda = 11$.

4.3. Results After Tuning Localization and Inflation

Once we have found the optimal values for localization and inflation, we can take a look at the results of assimilating the 51 observations with those parameters. These results are shown in Figure 10.

Panel (a) shows the results in observation space for the assimilation of the first observation, and it should be compared to panel (a) of Figure 6. In this case, the observation has a larger influence in the DA process, since the analysis values are closer to it.

Panel (b) shows the results in model space. To ease the visualization, we show the differences of the background and analysis profiles with respect to the reference. The background mean is shown as a blue dashed line, and it is accompanied by its standard deviation at each side of the mean, shown by the shaded cyan region. The 51 analysis means are shown with solid magenta lines. For the three variables, the profiles background and analysis means remain identical under 30 km. Above this height, the most benefited variable is the zonal wind. From 42 km up, the differences with respect to the reference are smaller for the analysis means than for the background mean. In particular, for the region between 42 and 50 km the analysis means have smaller differences with opposite sign from the background, and outside of the shaded cyan region. Note that changes above ∼45 km occur in a non-

21698996, 2026, 5, Downloaded from https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2025JD044841 by <Shibboleth>-member@reading.ac.uk, Wiley Online Library on [22/06/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

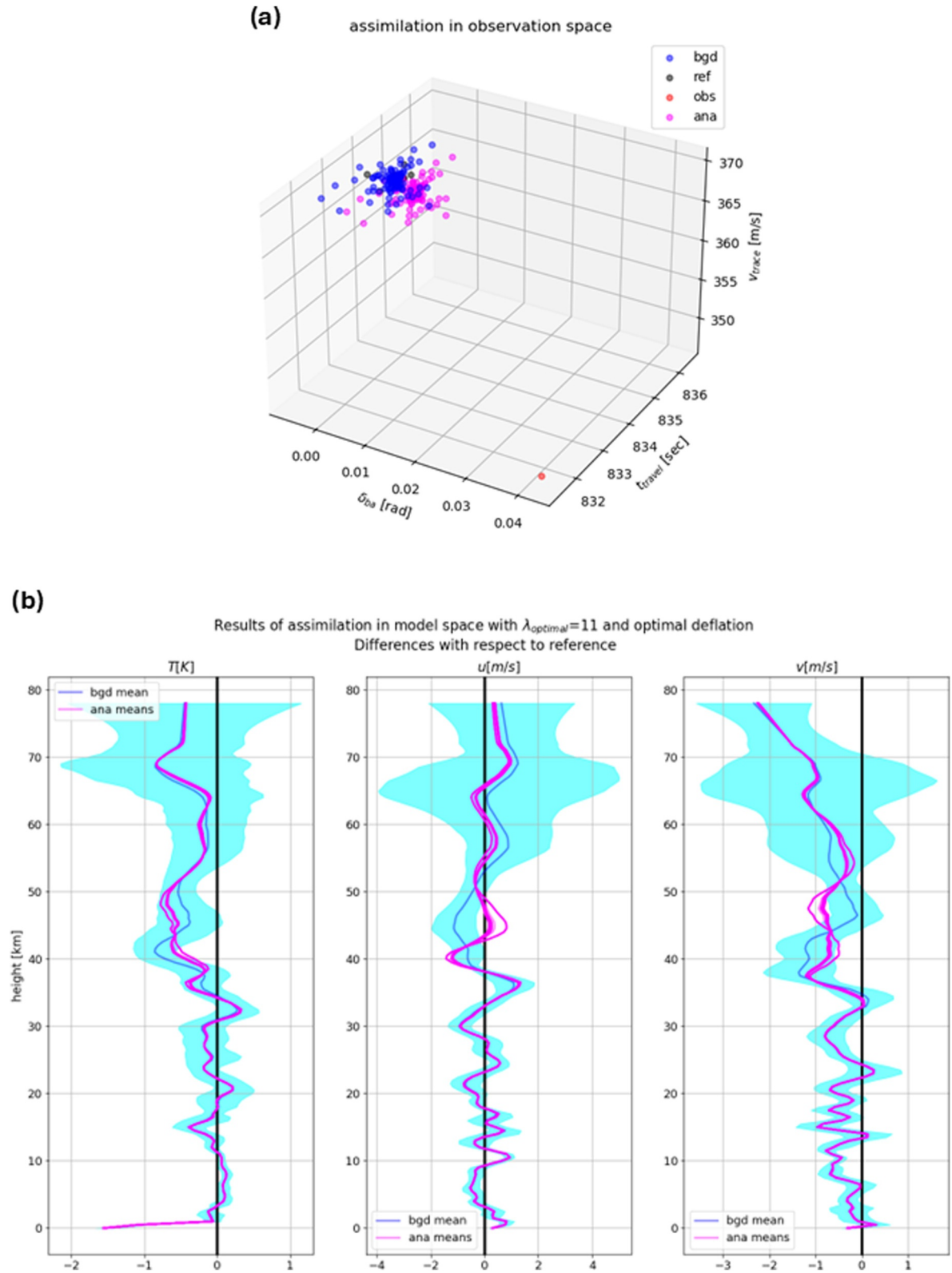

Figure 10. Results of assimilating infrasound observations with an optimal inflation and localization radius. (a) Results in observation space. (b) Results in model space plotted as differences with respect to the reference mean.

21698996, 2026, 5, Downloaded from https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2025JD044841 by <Shibboleth>-member@reading.ac.uk, Wiley Online Library on [22/06/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

observed region (since this region is above the refraction heights), so these changes come from the action of the covariances in the data assimilation process. For temperature and meridional wind the results are more mixed. The temperature error is reduced between 39 and 44 km, with some of this reduction being outside the shaded cyan region. However, there is an increase of the error between 44 and 51 km. For meridional wind, there is an decrease of the error between 38 and 45 km, as well as between 51 and 61 km. There is an increase of the error between 45 and 51 km.

Figure 11 shows the effects of the data assimilation on the effective sound speed. Once more, we compute the differences of the background and analysis means with respect to the references. As in the original state variables, there are no changes in the region below 34 km. Between 35 and 42 km, there analysis error is larger than the background error. However, above 42 km there is a consistent reduction of the error, and for the region between 42 and 50 km this reduction is outside the shaded cyan region.

In order to assess the impacts of the DA in more detail, it is useful to compute the so-called analysis increment for our different variables, which is defined as the difference between the analysis mean and the background mean:

$$\mathbf{d}^{ab} = \bar{\mathbf{x}}^{a} - \bar{\mathbf{x}}^{b}. \tag{22}$$

For the variables for which the assimilation of an observation had no impact, this increment is directly zero. Analysis increments (averaged over the 51 analysis means) are shown in Figure 12 for the three variables and each vertical level. Each panel is centered around zero. To assess the quality of these increments, we compute the following ratio for each variable:

$$r = \frac{|\bar{x}^{a} - \bar{x}^{\mathrm{ref}}|}{|\bar{x}^{b} - \bar{x}^{\mathrm{ref}}|}. \tag{23}$$

This is a ratio of errors. If the analysis mean is closer to the reference than the background mean, this ratio is smaller than one. On the contrary, a ratio greater than one indicates that the analysis is further from the reference than the background, a situation which is undesired but can still happen locally for some variables, even when globally we have shown that the MSE/DEN is lower for the analysis mean. We average this ratio over the 51 analysis means. In Figure 12 blue colors denote good increments (the darker the better), whereas red colors denote bad increments (the darker the worse).

We note that for the temperatures, the increments are quite small, but for the winds they are noticeable. In the case of zonal wind, there is a peak in increments of close to 1.5 m/s at around 45 km, and this is a good increment (under the assumption that the reference is indeed the best approximation of the true profile). Notice that there are increments above 50 and all the way up to 80 km with magnitude around 0.5 m/s. These are heights that the waves never reach, yet the impact of the observations is felt through the covariances as in the toy example. For the meridional wind, there is a peak increment of around 1.75 m/s at roughly 47 km, but it is a bad one. It is, however, accompanied by two good increments of around 0.5 m/s at 42 km, and 0.3 m/s at around 55 km. For the three variables, the impacts below 30 km are imperceptible.

We finally assess the impact of the DA in an indirect manner. When running the ray-tracing routine from a given atmospheric profile, a refraction height is diagnosed. In Figure 13 we display a violin plot comparing the refraction heights predicted for the background (left) and the center (reference). Then we compute ray-tracing on the analysis profiles and collect the predicted refraction heights. Notice that the heights coming from the analysis are lower than those of the background, and closer to those of the reference.

5. Summary and Discussion

Infrasound observations are routinely taken by different acoustic arrays around the world. As such, they are data sets of opportunity which can be exploited in a more thorough manner to detect atmospheric conditions.

This work has performed, for the first time, data assimilation on real infrasound observations while using a full-ray tracing routine as observation operator. This is an important improvement from previous works which used simulated observations and/or a simple geometric model as observation operator.

21698996, 2026, 5, Downloaded from https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2025JD044841 by <Shibboleth>-member@reading.ac.uk, Wiley Online Library on [22/06/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

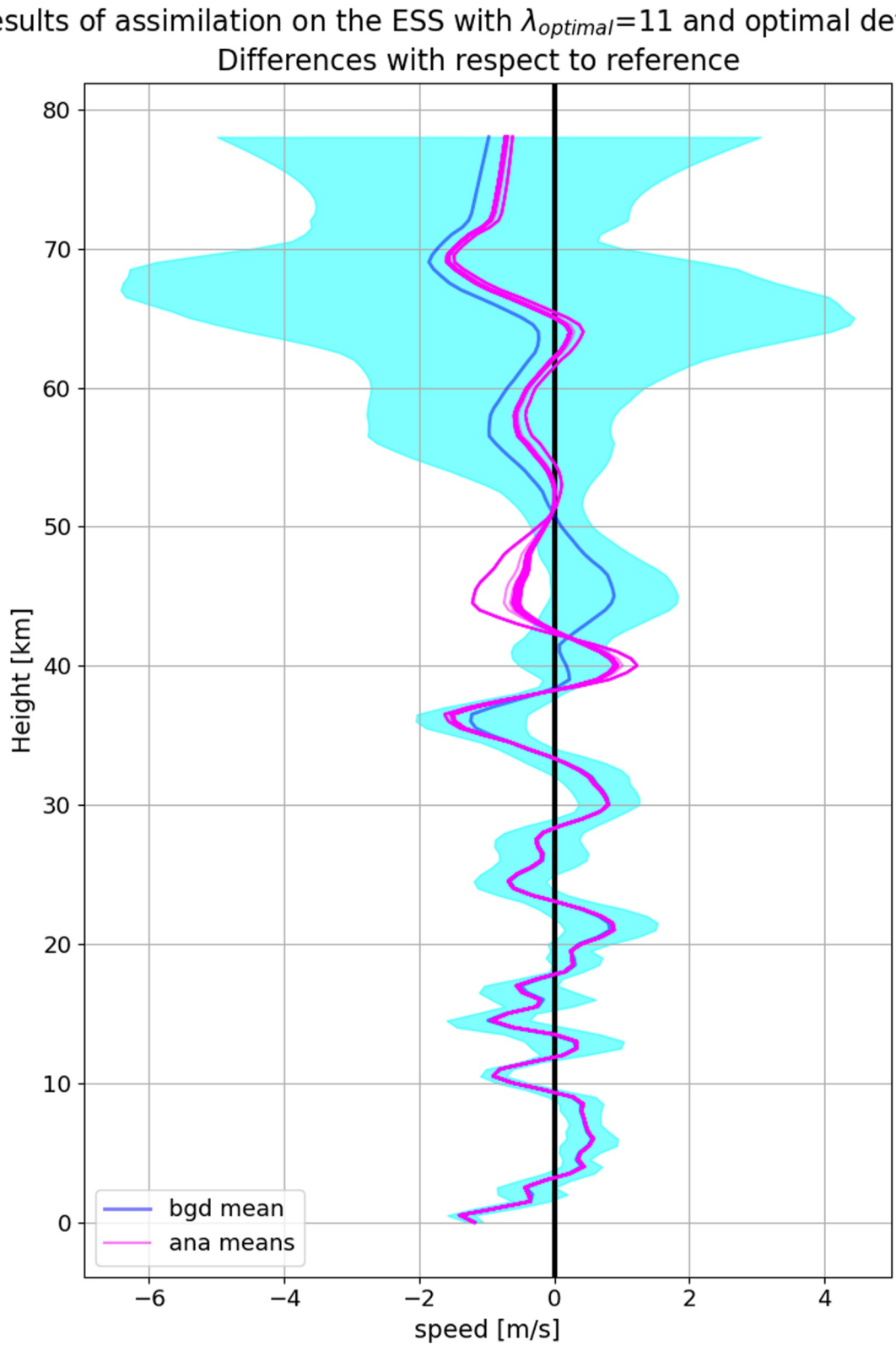

Figure 11. Effects of assimilating infrasound observations with an optimal inflation and localization radius on the effective sound speed. We have plotted the differences with respect to the reference mean.

21698996, 2026, 5, Downloaded from https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2025JD044841 by <Shibboleth>-member@reading.ac.uk, Wiley Online Library on [22/06/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

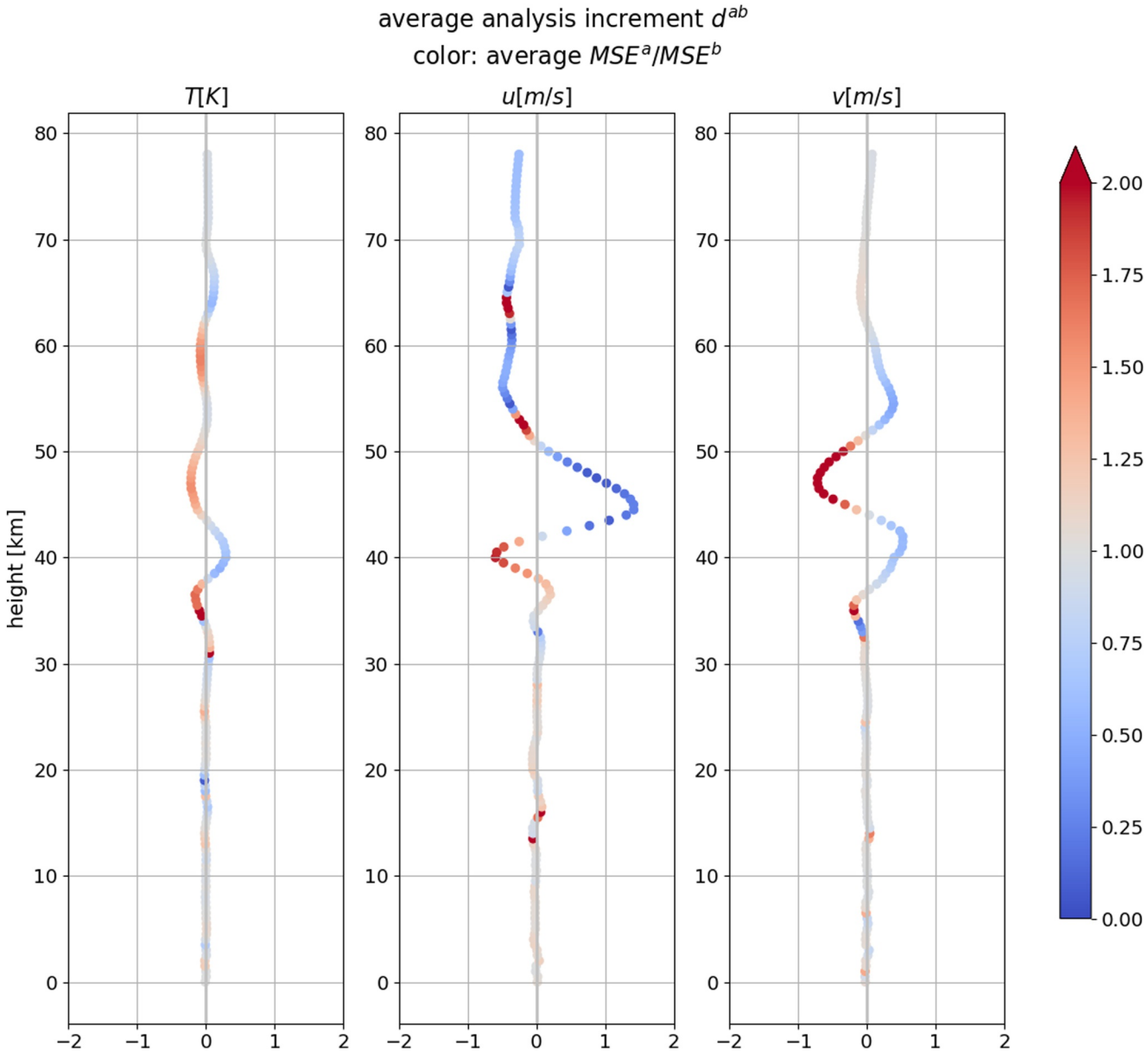

Figure 12. Analysis increments for our three variables (panels) at different levels. The coloring depends on the ratio of analysis MSE divided by background MSE.

Using the ensemble ERA5 product as background, we have shown that, after tuning the correct localization and inflation factors, infrasound observations can improve atmospheric profiles. These changes are more prominent at heights of 35–50 km, the range where the infrasound waves refract back to Earth. For temperatures, the changes are less than 0.5 K. However, for winds the changes can reach more than 2 m/s (especially for the zonal wind).

We have demonstrated that assimilating infrasound observations bring the atmospheric profiles closer to those obtained by solely interpolating the reanalysis product to the time of the explosions. We have shown some quantitative evidence of the improvement in the estimation of quantities in vertical columns. The ultimate aim, however, is to assimilate these novel observations in tandem with traditionally available observations like radiosondes. For instance, Francis et al. (2023) discuss the role of radiosondes as anchor observations in remote sensing. This can be done either in a regional or a global NWP model.

21698996, 2026, 5, Downloaded from https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2025JD044841 by <Shibboleth>-member@reading.ac.uk, Wiley Online Library on [22/06/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

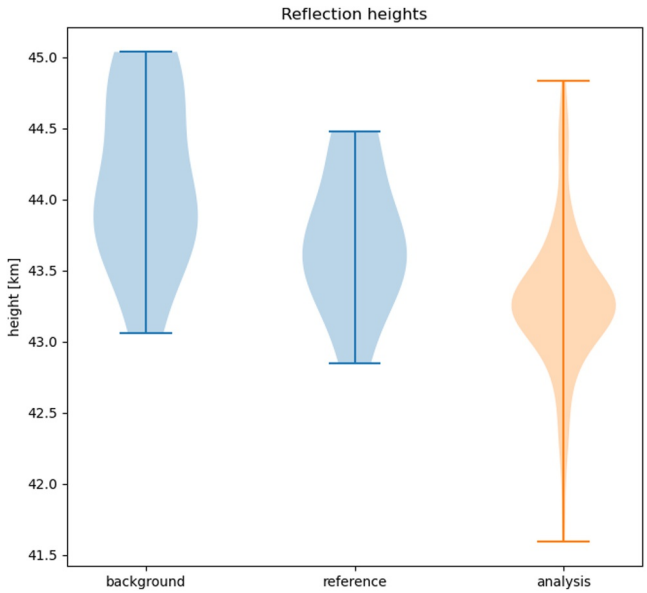

Figure 13. Refraction heights predicted by InfraGA for different profiles: background, reference and analysis.

While this paper emphasizes the use of infrasound to address observational gaps in the upper atmosphere, similar gaps exist in the troposphere where infrasound can also contribute. In particular, mesoscale meteorology is limited by the coarse spatial and temporal resolution of vertical wind measurements, especially within the boundary layer (Bell et al., 2020; Council et al., 2009; Mahmood et al., 2017). As in the middle atmosphere, infrasound waves can offer integrated observations of wind and temperature in the troposphere by propagating through shallow waveguides formed in the boundary layer or along the jet stream (G. Averbuch, Ronac-Giannone, et al., 2022). At these lower altitudes, a wider range of ambient infrasound sources are available, since smaller sources are audible at shorter distances.

Future areas of research include the following:

- The assimilation of data with various signal-to-noise ratios to estimate its effect. This will also include statistical analysis of the assimilation over multiple days of data.
- The assimilation of infrasound observations in tandem with anchor observations like those coming from radiosondes.
- The use of operational ensemble forecasts instead of the ERA5 reanalysis product. In this case, the benefit of assimilating infrasound observations may be larger since the forecasted profiles will be further away from the truth. We need to perform a systematic evaluation of the benefits as a function of the forecast lead time. There is a risk, however, in these experiments. Some atmospheric profiles represented in the ensemble forecast may not be so different from the real atmospheric profile that they do not have the conditions for the infrasound wave transmission. This was already pointed out in Amezcua et al. (2024) and we have yet to provide a solution for this problem.

The results obtained in this paper are encouraging toward the final goal of assimilating infrasound observations in a real NWP operational setting. A more challenging task will be to use infrasound observations from more ubiquitous sources, like microbaroms. In this continuous case, the source exact location and timing cannot be determined as with the detonations. Fortunately, some initial progress has been made on how to treat these observations, for example, Letournel et al. (2024). These ideas can be used a stepping stone toward the assimilation of these observations.

Conflict of Interest

The authors declare no conflicts of interest relevant to this study.

Data Availability Statement

The ERA5 reanalysis product is available in the Copernicus Climate Service (Hersbach et al., 2023). The infrasound raw data recorded from the Oklahoma field campaign can be downloaded from the Earth Scope Data Center (Averbuch & Arrowsmith, 2021).

Acknowledgments
The Authors thank two anonymous reviewers for their feedback that helped improve this paper. JA acknowledges the support of the University of Reading. This study was facilitated by previous research performed within the framework of the *Middle Atmosphere Dynamics: Exploiting Infrasound Using a Multidisciplinary Approach at High Latitudes* (MADEIRA) and *Assimilating infrasound data for multi-decade constraints on middle atmosphere wind and temperature* projects, funded by the Research Council of Norway under Grant Agreements 2537848 and 327313.

References

Amezcua, J., & Barton, Z. (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., Näsholm, S. P., Blixt, E. M., & Charlton-Perez, A. J. (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

Amezcua, J., Näsholm, S. P., & Rodriguez, I. V. (2024). Using satellite data assimilation techniques to combine infrasound observations and a full ray-tracing model to constrain stratospheric variables. *Monthly Weather Review*, *152*(8), 1883–1902. https://doi.org/10.1175/MWR-D-23-0186.1

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. (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., & Dey, C. H. (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

21698996, 2026, 5, Downloaded from https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2025JD044841 by <Shibboleth>-member@reading.ac.uk, Wiley Online Library on [22/06/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Asch, M., Bocquet, M., & Nodet, M. (2016). *Data assimilation*. Society for Industrial and Applied Mathematics. https://doi.org/10.1137/1.9781611974546

Assink, J., Smets, P., Marcillo, O., Weemstra, C., Lalande, J.-M., Waxler, R., & Evers, L. (2019). Advances in infrasonic remote sensing methods. In *Infrasound monitoring for atmospheric studies* (pp. 605–632). Springer. https://doi.org/10.1007/978-3-319-75140-5_18

Averbuch, G., Ronac-Giannone, M., Arrowsmith, S., & Anderson, J. F. (2022). Evidence for short temporal atmospheric variations observed by infrasonic signals: 1. The troposphere. *Earth and Space Science*, *9*(3), e2021EA002036. https://doi.org/10.1029/2021EA002036

Averbuch, G., Sabatini, R., & Arrowsmith, S. (2022). Evidence for short temporal atmospheric variations observed by infrasonic signals: 2. The stratosphere. *Earth and Space Science*, *9*(10), e2022EA002454. https://doi.org/10.1029/2022EA002454

Averbuch & Arrowsmith. (2021). Southern methodist university [Dataset]. *International Federation of Digital Seismograph Networks*. https://doi.org/10.7914/SN/3B_2021

Baker, W. E., Atlas, R., Cardinali, C., Clement, A., Emmitt, G. D., Gentry, B. M., et al. (2014). LiDAR-measured wind profiles: The missing link in the global observing system. *Bulletin of the American Meteorological Society*, *95*(4), 543–564. https://doi.org/10.1175/BAMS-D-12-00164.1

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

Bedard, A. J., Jr., & Georges, T. M. (2000). Atmospheric infrasound. *Physics Today*, *53*(3), 32–37. https://doi.org/10.1063/1.883019

Bell, T. M., Greene, B. R., Klein, P. M., Carney, M., & Chilson, P. B. (2020). Confronting the boundary layer data gap: Evaluating new and existing methodologies of probing the lower atmosphere. *Atmospheric Measurement Techniques*, *13*(7), 3855–3872. https://doi.org/10.5194/amt-13-3855-2020

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

Bishop, C. H., Whitaker, J. S., & Lei, L. (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., Ceranna, L., Hauchecorne, A., Charlton-Perez, A., Marchetti, E., Evers, L., et al. (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., Pol, K., Le Pichon, A., Hauchecorne, A., Keckhut, P., Baumgarten, G., et al. (2019). Middle atmosphere variability and model uncertainties as investigated in the framework of the ARISE project. In *Infrasound monitoring for atmospheric studies* (pp. 845–887). Springer. https://doi.org/10.1007/978-3-319-75140-5_28

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

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

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

Brissaud, Q., Martin, R., Garcia, R. F., & Komatitsch, D. (2016). Finite-difference numerical modelling of gravitoacoustic wave propagation in a windy and attenuating atmosphere. *Geophysical Journal International*, *206*(1), 308–327. https://doi.org/10.1093/gji/ggw121

Brissaud, Q., Näsholm, S. P., Turquet, A., & Le Pichon, A. (2022). Predicting infrasound transmission loss using deep learning. *Geophysical Journal International*, *232*(1), 274–286. https://doi.org/10.1093/gji/ggac307

Cameijo, A. J., Le Pichon, A., Sklab, Y., Arib, S., Brissaud, Q., Näsholm, S. P., et al. (2025). Towards real-time assessment of infrasound event detection capability using deep learning-based transmission loss estimation. https://doi.org/10.48550/arXiv.2506.06358

Campbell, W. F., Bishop, C. H., & Hodyss, D. (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

Charlton-Perez, A. J., Baldwin, M. P., Birner, T., Black, R. X., Butler, A. H., Calvo, N., et al. (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

Charlton-Perez, A. J., Bröcker, J., Karpechko, A. Y., Lee, S. H., Sigmond, M., & Simpson, I. R. (2021). A minimal model to diagnose the contribution of the stratosphere to tropospheric forecast skill. *Journal of Geophysical Research: Atmospheres*, *126*(24), e2021JD035504. https://doi.org/10.1029/2021JD035504

Council, N. R., on Earth, D., Studies, L., & on Atmospheric Sciences, B., & on Developing Mesoscale Meteorological Observational Capabilities to Meet Multiple National Needs, C. (2009). *Observing weather and climate from the ground up: A nationwide network of networks*. National Academies Press.

De Carlo, M., Hupe, P., Le Pichon, A., Ceranna, L., & Ardhuin, F. (2021). Global microbarom patterns: A first confirmation of the theory for source and propagation. *Geophysical Research Letters*, *48*(3), e2020GL090163. https://doi.org/10.1029/2020GL090163

Diamond, M. (1964). Cross wind effect on sound propagation. *Journal of Applied Meteorology*, *3*(2), 208–210. https://doi.org/10.1175/1520-0450(1964)003⟨0208:CWEOSP⟩2.0.CO;2

Domeisen, D. I. V., Butler, A. H., Charlton-Perez, A. J., Ayarzagüena, B., Baldwin, M. P., Dunn-Sigouin, E., et al. (2020a). The role of the stratosphere in subseasonal to seasonal prediction: 2. Predictability arising from stratosphere-troposphere coupling. *Journal of Geophysical Research: Atmospheres*, *125*(2), e2019JD030923. https://doi.org/10.1029/2019JD030923

Domeisen, D. I. V., Butler, A. H., Charlton-Perez, A. J., Ayarzagüena, B., Baldwin, M. P., Dunn-Sigouin, E., et al. (2020b). The role of the stratosphere in subseasonal to seasonal prediction: 1. Predictability of the stratosphere. *Journal of Geophysical Research: Atmospheres*, *125*(2), e2019JD030920. https://doi.org/10.1029/2019JD030920

Duruisseau, F., Huret, N., Andral, A., & Camy-Peyret, C. (2017). Assessment of the ERA-Interim winds using high-altitude stratospheric balloons. *Journal of the Atmospheric Sciences*, *74*(6), 2065–2080. https://doi.org/10.1175/JAS-D-16-0137.1

Evensen, G. (2006). *Data assimilation: The ensemble Kalman filter*. Springer-Verlag. https://doi.org/10.1007/978-3-642-03711-5

Farchi, A., & Bocquet, M. (2019). On the efficiency of covariance localisation of the ensemble Kalman filter using augmented ensembles. *Frontiers in Applied Mathematics and Statistics*, *5*, 3. https://doi.org/10.3389/fams.2019.00003

21698996, 2026, 5, Downloaded from https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2025JD044841 by <Shibboleth>-member@reading.ac.uk, Wiley Online Library on [22/06/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Francis, D. J., Fowler, A. M., Lawless, A. S., Eyre, J., & Migliorini, S. (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/10.1002/qj.4482

Gaspari, G., & Cohn, S. E. (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

Gibbons, S., Kværna, T., & Näsholm, P. (2019). Characterization of the infrasonic wavefield from repeating seismo-acoustic events. In *Infrasound monitoring for atmospheric studies* (pp. 387–407). Springer. https://doi.org/10.1007/978-3-319-75140-5_10

Greybush, S. J., Kalnay, E., Miyoshi, T., Ide, K., & Hunt, B. R. (2011). Balance and ensemble Kalman filter localization techniques. *Monthly Weather Review*, *139*(2), 511–522. https://doi.org/10.1175/2010MWR3328.1

Hamill, T. M., Whitaker, J. S., & Snyder, C. (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., Bell, B., Berrisford, P., Biavati, G., Horányi, A., Muñoz Sabater, J., et al. (2023). ERA5 hourly data on single levels from 1940 to present [Dataset]. *Copernicus Climate Change Service (C3S) Climate Data Store (CDS)*. https://doi.org/10.24381/cds.adbb2d47

Hersbach, H., Bell, B., Berrisford, P., Horányi, A., Muñoz Sabater, J., Nicolas, J., et al. (2019). Global reanalysis: Goodbye ERA-Interim, hello ERA5. *ECMWF Newsletter*, *159*, 17–24. https://doi.org/10.21957/vf291hehd7

Houtekamer, P. L., & Zhang, F. (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

Jazwinski, A. (1970). *Stochastic processes and filtering theory (No. 64)*. Acad. Press.

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

Kalman, R. E., & Bucy, R. S. (1961). New results in linear filtering and prediction theory. *Transactions of the ASME. Series D, Journal of Basic Engineering*, *109*.

Kalnay, E. (2003). *Atmospheric modeling, data assimilation and predictability*. Cambridge University Press. https://doi.org/10.1017/CBO9780511802270

Kidston, J., Scaife, A. A., Hardiman, S. C., Mitchell, D. M., Butchart, N., Baldwin, M. P., & Gray, L. J. (2015). Stratospheric influence on tropospheric jet streams, storm tracks and surface weather. *Nature Geoscience*, *8*(6), 433–440. https://doi.org/10.1038/ngeo2424

Korhonen, N., Hyvärinen, O., Kämäräinen, M., Richardson, D. S., Järvinen, H., & Gregow, H. (2019). Adding value to extended-range forecasts in northern Europe by statistical post-processing using stratospheric observations. *Atmospheric Chemistry and Physics Discussions*, *2019*, 1–18. https://doi.org/10.5194/acp-2019-679

Le Pichon, A., Blanc, E., & Hauchecorne, A. (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

Letournel, P., Listowski, C., Bocquet, M., Le Pichon, A., & Farchi, A. (2024). Evaluating numerical weather prediction models in the middle atmosphere using coherent oceanic acoustic noise observations. *Journal of Geophysical Research: Atmospheres*, *129*(23), e2024JD042034. https://doi.org/10.1029/2024JD042034

Mahmood, R., Boyles, R., Brinson, K., Fiebrich, C., Foster, S., Hubbard, K., et al. (2017). Mesonets: Mesoscale weather and climate observations for the United States. *Bulletin of the American Meteorological Society*, *98*(7), 1349–1361. https://doi.org/10.1175/BAMS-D-15-00258.1

Marty, J. (2019). The IMS infrasound network: Current status and technological developments. In *Infrasound monitoring for atmospheric studies* (pp. 3–62). Springer.

Mitchell, D. M., Gray, L. J., Anstey, J., Baldwin, M. P., & Charlton-Perez, A. J. (2013). The influence of stratospheric vortex displacements and splits on surface climate. *Journal of Climate*, *26*(8), 2668–2682. https://doi.org/10.1175/JCLI-D-12-00030.1

Nerger, L., Janjić, T., Schröter, J., & Hiller, W. (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

Orsolini, Y. J., Kindem, I. T., & Kvamstø, N. G. (2011). On the potential impact of the stratosphere upon seasonal dynamical hindcasts of the North Atlantic Oscillation: A pilot study. *Climate Dynamics*, *36*(3–4), 579–588. https://doi.org/10.1007/s00382-009-0705-6

Polavarapu, S., Shepherd, T., Rochon, Y., & Ren, S. (2005). Some challenges of middle atmosphere data assimilation. *Quarterly Journal of the Royal Meteorological Society*, *131*(613), 3513–3527. https://doi.org/10.1256/qj.05.87

Rennie, M. P., Isaksen, L., Weiler, F., de Kloe, J., Kanitz, T., & Reitebuch, O. (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

Scaife, A. A., Baldwin, M. P., Butler, A. H., Charlton-Perez, A. J., Domeisen, D. I. V., Garfinkel, C. I., et al. (2022). Long-range prediction and the stratosphere. *Atmospheric Chemistry and Physics*, *22*(4), 2601–2623. https://doi.org/10.5194/acp-22-2601-2022

Stuart, A. M. (2010). Inverse problems: A Bayesian perspective. *Acta Numerica*, *19*, 451–559. https://doi.org/10.1017/S0962492910000061

Talagrand, O. (1981). On the mathematics of data assimilation. *Tellus*, *33*(4), 321–339. https://doi.org/10.1111/j.2153-3490.1981.tb01755.x

Tippett, M. K., Anderson, J. L., Bishop, C. H., Hamill, T. M., & Whitaker, J. S. (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

Vorobeva, E., De Carlo, M., Le Pichon, A., Espy, P. J., & Näsholm, S. P. (2021). Benchmarking microbarom radiation and propagation model against infrasound recordings: A vespagram-based approach. *Annales Geophysicae*, *39*(3), 515–531. https://doi.org/10.5194/angeo-39-515-2021

Vorobeva, E., Eggen, M. D., Midtfjord, A. D., Benth, F. E., Hupe, P., Brissaud, Q., et al. (2024). Estimating stratospheric polar vortex strength using ambient ocean-generated infrasound and stochastics-based machine learning. *Quarterly Journal of the Royal Meteorological Society*, *150*(762), 2712–2726. https://doi.org/10.1002/qj.4731

Wang, X., Bishop, C. H., & Julier, S. J. (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

Waugh, D. W., Sobel, A. H., & Polvani, L. M. (2017). What is the polar vortex and how does it influence weather? *Bulletin of the American Meteorological Society*, *98*(1), 37–44. https://doi.org/10.1175/BAMS-D-15-00212.1

21698996, 2026, 5, Downloaded from https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2025JD044841 by <Shibboleth>-member@reading.ac.uk, Wiley Online Library on [22/06/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License