Submitted November 14, 2025

ORIGINAL ARTICLE

EnKF and EM based parameter estimation of a convective gravity wave parameterization using Strateole 2 constant level balloon data

F. Lott[1*] | **P. Tandeo**[2†] | **M. Pulido**[3†] | **D. Bardet**[1‡]

[1]Laboratoire de Météorologie Dynamique, Sorbonne-Universite, PSL/Ecole Normale Supérieure, Paris France

[2]Odyssey, Inria/IMT/CNRS, F-29280 Plouzané, France; IMT Atlantique, Lab-STICC, UMR CNRS 6285, 29238, Brest, France; RIKEN Cluster for Pioneering Research, Kobe, 650-0047, Japan

[3]Departamento de Física, Facultad Ciencias Exactas y Naturales y Agrimensura, Universidad Nacional del Nordeste, Corrientes, Argentina

Correspondence
Email: francois.lott@sorbonne-universite.fr

Funding information
VESRI Schmidt Future project DataWave

An offline methodology is applied to estimate parameters of a subgrid-scale non-orographic gravity-wave scheme using observations from constant-level balloons. The approach integrates the Ensemble Kalman Filter (EnKF) with an iterative parameter estimation method based on the expectation-maximization (EM) algorithm. The meteorological fields required for the parameterization offline are taken from the ERA5 reanalysis, corresponding to the instantaneous meteorological conditions found underneath the Strateole-2 balloon observations made in the lower tropical stratosphere from November 2019 to February 2021 and October 2021 to January 2022. Compared to a direct approach that minimizes a cost function and uses Bayesian inference of parameters, our analysis demonstrates that the EnKF/EM method effectively characterizes the launching amplitudes and altitudes of the parameterized gravity waves and while quantifying their associated uncertainties. Furthermore, we illustrate how the method can help improving a scheme, specifically the results indicate that introducing a background wave activity renders the convective wave parameterization more

[*] Conceptualization, Formal analysis, Methodology, Writing
[†] Formal analysis, Methodology, Writing
[‡] Numerical Analysis

realistic.

KEYWORDS
Non-orographic gravity waves, parameterization, Ensemble Kalman filter, Constant level balloon observations, Parameter estimation, Uncertainty quantification

1 | INTRODUCTION

It is well known that large-scale circulation in the middle atmosphere is largely driven by gravity waves (GWs) that transport horizontal momentum and modify large-scale flow when they break (Andrews et al., 1987). In the tropics, for instance, non-orographic gravity waves contribute significantly to the forcing of the Quasi-Biennial Oscillation (QBO), a nearly 28-month oscillation of the zonal mean zonal winds that occurs in the lower part of the equatorial stratosphere (Baldwin et al., 2001). These waves have horizontal scales much shorter than the 100 km typical horizontal resolution of climate global models, and need to be parameterized. Accordingly, almost all state of the art global climate models that simulate a QBO include non-orographic GWs parameterization schemes (Scinocca, 2003; Song and Chun, 2005; Beres et al., 2005; Orr et al., 2010; Lott and Guez, 2013; Bushell et al., 2015; Anstey et al., 2016; Christiansen et al., 2016; Serva et al., 2018). In these schemes, the parameters controlling GW forcing are tuned so that large-scale model features, such as the period and amplitude of the QBO are close to the observed ones.

Although GWs parameterizations are now used in models successfully, their validation using direct in situ observations remains a challenge. Large horizontal-scale GWs can be obtained from global satellite observations of temperature (Geller et al., 2013) and the corresponding momentum flux computed using polarization relations (Alexander et al., 2010; Ern et al., 2014). However, in situ observations are required to observe the shorter horizontal scales that also contribute to the QBO forcing and to obtain a direct measurement of the corresponding momentum flux. The most precise measurements are provided by constant-level long-duration balloons, like those made in the deep tropics during PreConcordiasi (Jewtoukoff et al., 2013), Strateole 2 (Haase et al., 2018) and some of the LOON commercial flights (Green et al., 2023). Among the important results, these balloon observations have shown that the momentum flux entering the stratosphere is extremely intermittent (Hertzog et al., 2012). This intermittency implies that the mean momentum flux is mostly transported by few large-amplitude GWs that are expected to break at lower altitudes than when it is transported by more uniform and smaller amplitude GWs. Also, Lott et al. (2023) and Green et al. (2023) have shown, using Strateole-2 and Loon data respectively, that this intermittent character can be characterized by log-normal distributions of the momentum fluxes. Lott et al. (2023) also showed that the MF evaluated offline using LMDz parameterizations and ERA5 data colocated with the balloon locations provide MF estimates comparable in amplitude to those needed in this GCM to simulate a QBO.

In the modeling community, it is well established that the tuning of the QBO is sensitive to GW scheme parameters, with all existing schemes having parameters directly related to the launched waves amplitude, phase speed distribution, launching altitude, and breaking detection. These parameters control the modeled QBO amplitude and periods (Giorgetta et al., 2006; Geller et al., 2016; Garfinkel et al., 2022), but are poorly constrained by observations. For instance, in Garfinkel et al. (2022) varying the amplitude of the launch GWs by a factor near 5 does not significantly affect the QBO period, whereas in Mansfield and Sheshadri (2024) the acceptable range of values is significantly narrower. In Mansfield and Sheshadri (2024) the evaluation of the parameters and their uncertainties is conducted by comparing model responses to large-scale observed QBO statistics. As in Annan et al. (2005); Dunbar et al. (2021);

Cleary et al. (2021) they combine ensemble Kalman filters and Bayesian estimations, techniques that can also include ensembles of high-resolution simulations of the parameterized processes (Couvreux et al., 2021). Despite these indirect approaches, the direct use of gravity wave observations to tune parameterizations in an offline mode remains, to our knowledge, limited. These considerations lead us to question whether the typically "chosen" parameter values are consistent with direct measurements, and whether the parameter sensitivity observed in models reflects underlying uncertainties in both the parameterizations and the observations.

For gravity waves, some benefits could arise since parameterizations in offline modes have skill in predicting aspects of the subgrid scale processes and are quite realistic in reproducing their online behavior (Lott et al., 2023). Nevertheless, we can expect difficulties because the GW parameterizations used in offline modes driven by data are still subject to substantial and nearly unquantifiable uncertainties. For instance, the sampling of the gravity wave scheme by a balloon only concerns single trajectories and not an entire model grid cell, the collocation of the balloon and the meteorological fields is only approximate. We can also question the technique used to characterize the evaluation of momentum fluxes from observations, despite the fact that the measurement errors are not perceived to be large. Finally, the parameterizations are oversimplified; for instance, they often assume constant launching altitudes, a uniform breaking criterion, simplifying relations with potential sources, if any, and neglect lateral propagation (Achatz et al., 2023; Voelker et al., 2024). Their strength is that they are based on a few physically understood mechanisms, like the Eliasen-Palm theorem or the WKB theory of monochromatic gravity wave vertical propagation.

In the general problem of parameter estimation using direct observations and a forecast model, conventional techniques consist of using the Ensemble Kalman Filter (EnKF) and Ensemble Kalman Smoother (EnKS), which are reviewed in Evensen (2009) and Carrassi et al. (2018). The classic way is to use an augmented state formulation, allowing the ensemble Kalman technique to identify correlations between the state of the system and the physical parameters (Ruiz et al., 2013; Ruiz and Pulido, 2015). However, this approach is computationally demanding because the model must be evaluated at each time step for each ensemble member. Moreover, tuning the model error covariance is nontrivial and requires additional estimation techniques Tandeo et al. (2020).

In this article, as the parameterizations are estimated offline we cannot use augmented states of a GCMs to evaluate them. We therefore follow Tandeo et al. (2015) by considering that the physical parameters are the only state variables that evolve. These parameters are assumed to follow a random walk, thereby incorporating non-negligible errors. This technique infers the parameter statistics using an iterative maximum likelihood approach based on the Expectation Maximization (EM) algorithm introduced by Dempster et al. (1977). In Tandeo et al. (2015) the method was found to behave well in a synthetic setup, using the Lott and Miller (1997) subgrid scale orography parameterization scheme, assuming observations consisted of 30 days of parameterized drag with given values of the scheme parameters. In this synthetic setup, the method converges and recovers the parameters. The purpose of the present paper is to examine the methodology by applying it to a nonorographic gravity wave parameterization in a realistic setup, e.g. using real observations containing uncertainties and assuming that the parameters can vary over time for physical reasons. A key outcome of the method is the quantification of parameter uncertainties. To evaluate the output of the technique, we compare the results of the EnKF/EM estimation with a direct Bayesian estimation. The latter is based on probabilities built from kernel density estimates of the observations and model errors.

The plan of the paper is as follows, section 2 recalls the basic aspects of the gravity wave parameterization used, and Section 3 provides a first "direct" appreciation of its sensitivity to parameter changes. As some parameters do not produce much variation in GWs momentum fluxes and others affect GWs dynamics far aloft the balloon altitude, we next select the parameters that directly control the launched momentum flux altitude and amplitude. In Section 4 we present the EnKF/EM method applied to our data. Section 5 discusses the results from the EnKF/EM and compares its outcomes to those of the direct Bayesian estimation. Section 6 summarizes the result, discuss the limits and potential

future applications.

2 | PARAMETERIZATION AND DATA USED

The parameterization scheme we use follows the convective GWD parameterization scheme presented in Lott et al. (2023) and that is operational in the IPSLCM6 climate model (Boucher et al., 2020). The only change we test here is the introduction of a background wave activity detailed below. In short, it is a multiwave stochastic scheme that represents the GW field by a stochastic sum of monochromatic waves in the zonal direction (x) and time (t) at each physical time step and at each horizontal gridpoint. The induced vertical velocity by these monochromatic waves is

$$w' = \sum_{j=1}^{J} C_j \mathbf{w}_j(z) e^{z/2H} e^{i\left(k_j x - \omega_j t\right)}, \tag{1}$$

where x points in the zonal direction, z is the log pressure coordinate, and $H = 7$ km is a characteristic scale height. The vertical variations of the complex vertical velocity amplitude $\mathbf{w}_j(z)$ vary according to the WKB approximation. The induced horizontal velocity is derived from the local polarization relation. The intermittency coefficients, C_j, are chosen to satisfy $\sum_{j=1}^{J} C_j^2 = 1$. They could be chosen randomly, but we fixed them at $C_j = J^{-1/2}$. The sign of horizontal wavenumbers determines the direction of propagation with respect to the zonal wind at the launch level, with $k_j > 0$ and k_j<0 for eastward and westward propagating waves, respectively. Their amplitudes are constrained to the range,

$$k_{\text{min}} < |k_j| < k_{\text{max}}, \tag{2}$$

where k_{min} is the horizontal wavenumber below which we consider that the GWs cannot be explicitly resolved by the model. It is a tunable parameter but we leave it fixed in the following at the value $k_{\text{min}} = 2\pi/(300\text{km})$, assuming that 300km is around the minimum horizontal scale a CMIP-type GCM with 1^o horizontal resolution can fully resolve. We take $k_{\text{max}} = 2\pi/(6\text{km})$ which is a conservative upper bound for vertically propagating hydrostatic gravity waves. We choose to sample the intrinsic zonal phase speed (c_j) at the launching altitude (z_l) according to a normal distribution

$$c_j(z_l) \sim \mathcal{N}(0, C_{\text{max}}), \tag{3}$$

where the characteristic phase speed C_{max} is a tunable parameter and U is the gridscale zonal wind. Once the intrinsic zonal phase speed is selected, the (absolute) frequencies are then determined $\omega_j = k_j\,[c_j(z_l) + U(z_l)]$, which are taken as positive by convention.

In Lott and Guez (2013), it is described how gridscale precipitation P results in GWs, we use the corresponding expression for the launched momentum fluxes and add to it a constant background:

$$F_j(z_l) = \Re\left\{\rho \mathbf{u}_j(z_l)\mathbf{w}_j^*(z_l)\right\} = \frac{k_j}{|k_j|}\left(G_{uw0}\left(\frac{k_j R L_c P}{H c_p}\right)^2 \frac{e^{-m_j^2 \Delta z^2}}{\rho_r N \Omega_j^3} + G_{uwb}\right). \tag{4}$$

where $\Omega_j = \omega_j - k_j U$ is the intrinsic frequency, $\rho = \rho_r e^{-z/H}$ and R, L_c and C_p are the dry air constant, the latent heat of condensation and the heat capacity at constant pressure, $m_j = \frac{N}{U - c_j}$ is the vertical wavenumber amplitude, and N is the buoyancy frequency. The tuning parameters Δz, G_{uw0} and G_{uwb} represent the heating source depth, the convective wave amplitude and the amplitude of background wave activity, respectively.

The breaking criterion can be expressed when passing from one model level z to the next level above $z + \delta z$ by writing it as

$$F_j(z+\delta z) = k_j \Theta\left(c_j(z+\delta z)c_j(z)\right) \mathrm{Min}\left\{ \frac{F_j(z)}{k_j} e^{-2\frac{\mu m^3}{\rho\Omega}\delta z}, \rho S^2 \frac{k_{\mathrm{min}}^2}{N} \frac{c_j^3}{k_j^2} \right\} \tag{5}$$

where the Heaviside function Θ represents wave absorption by critical levels. The first term in the parentheses expresses the decay of the EP-flux due to diffusion, μ. The last term is the saturated EP-flux and S is the saturation parameter. As the intrinsic phase speed c_j depends on the changes in the background wind (see (3)), this term is also referred to as "dynamical filtering". It is strongly affected by the choice of C_{max} as well.

The fluxes in (5) are computed using hourly data of zonal wind, temperature and precipitation from ERA5 at the balloon horizontal location. They are separated in the eastward component and the westward component by selecting the harmonics with $k_j > 0$ and $k_j < 0$ and by summing them separately. In the following, we will also average the results over time, the minimum length is one day. Then, the parameterized momentum flux vector is

$$\left(F^{fE}, F^{fW}\right) = \left(\left\langle \sum_{j=1}^{J} C_j^2 \Theta(k_j) F_j(z_b) \right\rangle, \left\langle \sum_{j=1}^{J} C_j^2 \Theta(-k_j) F_j(z_b) \right\rangle \right), \tag{6}$$

where z_b is the balloon flight altitude, the brackets indicate daily averages, and the Θ function applied to the zonal wavenumber sign selects the eastward and westward contributions to the momentum fluxes. In what follows, we will use a more concise notation and write the parameterized momentum flux vector,

$$\left(\boldsymbol{F}^{f}(t_k, \boldsymbol{\theta})\right)^{\top} = \left(F^{fE}, F^{fW}\right)[\boldsymbol{Z}(t_k), \boldsymbol{\theta}]\,. \tag{7}$$

where the $\boldsymbol{Z}(t_k)$ are hourly inputs from ERA5 and the parameters $\boldsymbol{\theta}$ are the ones to be estimated. Therefore, $\boldsymbol{F}^{f\top}(t_k)$ represents the forecasted or estimated momentum flux by the parameterization at time t_k given a set of parameters $\boldsymbol{\theta}$ and inputs $\boldsymbol{Z}$. The superscript f means forecast corresponding to the standard notation used in data assimilation.

The observations we use are derived from the instantaneous values of momentum fluxes in the Eastward and Westward directions, provided every 30s and derived through wavelet analysis (Corcos et al., 2021). The method permits the separation of disturbances with periods between 15min and 1h and disturbances with periods between 15min and 1day, but we follow Lott et al. (2023) and assume that only the fastest one needs to be used in order to tune a GW parameterization. The Eastward and Westward fluxes are separated at an early stage, each is then averaged over a day, in part because it takes about a day for a balloon to fly across a typical GCM gridbox. This defines the observation vector,

$$\boldsymbol{F}^{oT}(t_k) = \left(F^{oE}, F^{oW}\right)[t_k]\,, \tag{8}$$

for the $K = 1294$ number of days of Strateole 2 balloon measurements at altitudes near $z = 20$km. This number of days is the sum of days during which the 23 balloons flew, taking into account the 23 balloon flight during Strateole phase 1 and 2 that last more than 1 week (some balloons flew up to near two months).

At this stage, the parameterization potentially has 9 tuning parameters, among which 4 are directly related to the amplitude and location of the gravity wave sources in the troposphere, G_{uw0}, G_{uwb}, z_l and Δz , while k_{min}, k_{max}, C_{max} are associated with the characteristics of the parameterized waves. Two parameters μ and S control gravity

wave absorption by critical levels via wave dissipation and saturation. As we only have a limited sample of independent data, and because these data are collected at one level not all the parameters are identifiable because they are not independent. Preliminary systematic sensitivity evaluations show that the results at the balloon level are slightly sensitive to the choices in $k_{\min}$ and $k_{\max}$ when varied within reasonable bounds (multiplied or divided by a factor of 2). Furthermore, they are also almost insensitive to the dissipation μ that is used to dissipate waves near the model top (around $z = 80$km). Therefore, in what follows, we disregard these 3 parameters from the analysis and choose to estimate six parameters from observations

$$\boldsymbol{\theta}^T = (G_{uw0}, S, p_l, \Delta z, C_{\max}, G_b), \tag{9}$$

where $p_l = p_s e^{-z_l/H}$ is the pressure at the launching level and $p_s = 1$bar.

3 | DIRECT METHOD

To analyze how the outcome of the GW scheme at the balloon level varies with the six selected parameters in (9), we evaluate the cost function under K background conditions,

$$\mathbf{J}(\boldsymbol{\theta}) = \frac{1}{K}\sum_{k=1}^{K}\left\{\boldsymbol{F}^f(t_k,\boldsymbol{\theta}) - \boldsymbol{F}^o(t_k)\right\}^T\left\{\boldsymbol{F}^f(t_k,\boldsymbol{\theta}) - \boldsymbol{F}^o(t_k)\right\} \tag{10}$$

and vary these parameters randomly in the hypercube

$$[0.03, 3] \times [0.03, 3] \times [0.1, 1\text{bar}] \times [5\text{m}, 5\text{km}] \times [1, 300\text{m/s}] \times [0.1\text{mPa}, 10\text{mPa}]$$

with a uniform distribution for p_l and uniform distributions for the logarithm of the other variables. We then generate 8^6 realizations of the parameters, which are used in the scheme in combination with all the Strateole 2 data ($K = 1294$days). From this ensemble of 8^6 simulations, we found that the minimum of the cost function occurs near

$$\theta_m = (G_{uw0} = 0.15, S = 0.3, DZ = 100\text{m}, p_l = 0.45\text{bar}, C_{\max} = 40\text{m/s}. G_b = 1\text{mPa}) \,. \tag{11}$$

Figure 1a) shows time series of eastward and westward MFs measured during all balloon flights and the corresponding offline estimates using the optimal parameters and the operational ones. Figure 1b) is a zoomed view of flight 2 from phase 1. The time axis in these panels (in days) is somewhat artificial, as multiple balloons were flying concurrently. To create a continuous timeline, the time since launch for each balloon is appended to the total duration of the previously launched balloon. Overall, the amplitudes of the parameterized momentum fluxes for the 1-hour to 15-minute periods agree reasonably well with the measurements. A key difference, however, is evident in the operational parameterization: without a background term, the fluxes are often too small during periods of low precipitation. This is compensated for by peak fluxes that frequently exceed the observed values. In contrast, the simulation with optimal parameters more accurately captures the low-amplitude fluxes and produces less pronounced peaks.

The parameterization estimates fluxes of about the right amplitude, as shown in Fig. 1c), in which the average of the fluxes over the 23 entire flights are shown. It confirms systematically that the offline estimations are quite good on average with significant improvement when the optimal values are chosen. The parameterized fluxes are much more aligned with the measured ones. This improvement is even more pronounced when looking at the net

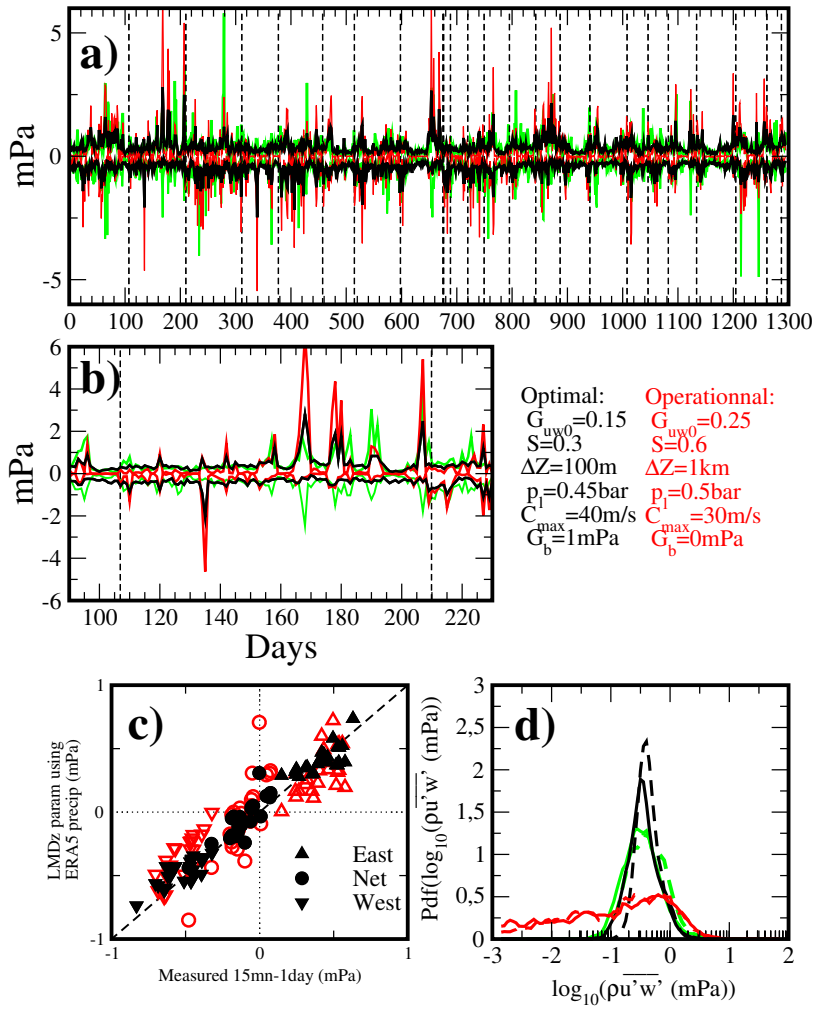

FIGURE 1 a) Daily sequence of MFs along the Strateole 2 campaign; b) Zoom over the 2d flight. c) Scatterplots of parameterized versus measured MF averaged over the Strateole2 balloon flights. d) PDFs of the daily MFs with eastward MF in solid and westward in dash. In black, the parameter values are near those yielding minima in cost function (see Eq. 11 and the crosses in Fig. 2). In red are the series for the operational values of the parameters and in green the observations.

fluxes (the circles in Figs. 1). The curves in Figs. 1a)-1b) also suggest that observations and offline estimations evolve quite similarly day after day, with the operational values being much more intermittent than the optimal ones. On the other hand, the optimal values underestimate the peak values. Figs. 1d) present probability density functions (PDFs) of the distributions of the momentum fluxes considering all the daily data. The balloons almost systematically measure fluxes with amplitudes between 0.1 and 1mPa (green curves), whereas in the operational parameterization there are many more contributions from the smaller amplitude momentum fluxes (red), not to mention that the zero values are excluded from PDFs when plotted versus the logarithm of MF amplitudes. With the addition of a constant background and for the optimal values (11) the black curves show that the distribution of MFs is more similar to the observed one. The PDFs are, nevertheless, substantially narrower, which means that the parameterized fluxes become much less intermittent. The PDFs are plotted using logarithmic scales, illustrating that the observed PDFs in green are reasonably log-normal. This property is essential in GWs parameterization because it translates the intermittency of the flux (Green et al., 2023) with broader distributions corresponding to more intermittent fluxes.

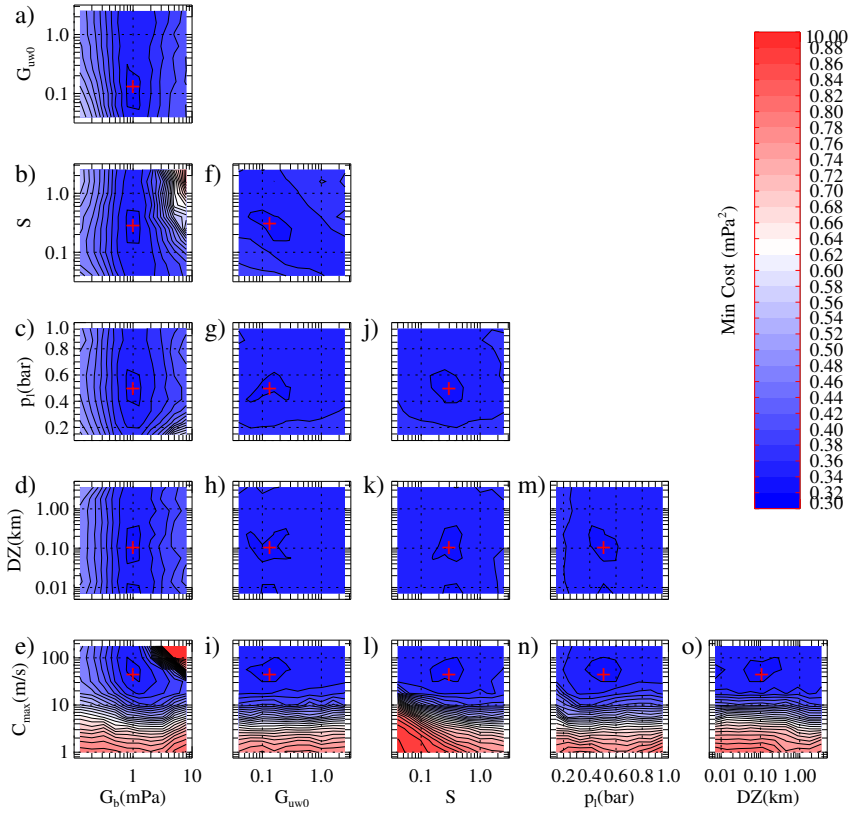

FIGURE 2 The sensitivity of the cost function when the six parameters G_{uw0}, G_b, S, Δz, z_l, and $C_{\max}$ are varied randomly with 8^6 realisations. In each panel, the cross-section of the the cost function is shown when the two values of the parameters shown in the x and y axis are varied and the rest is kept fixed. Red crosses indicate parameter given in (11) that gives the minima of (Cost = 0.34mPa^2) are given.

To appreciate the sensitivity of the cost function to individual parameter changes, each panel in Fig. 2 shows the minima of the cost when the 4 parameters not shown on the axes vary. In panel (a), the cost function does not

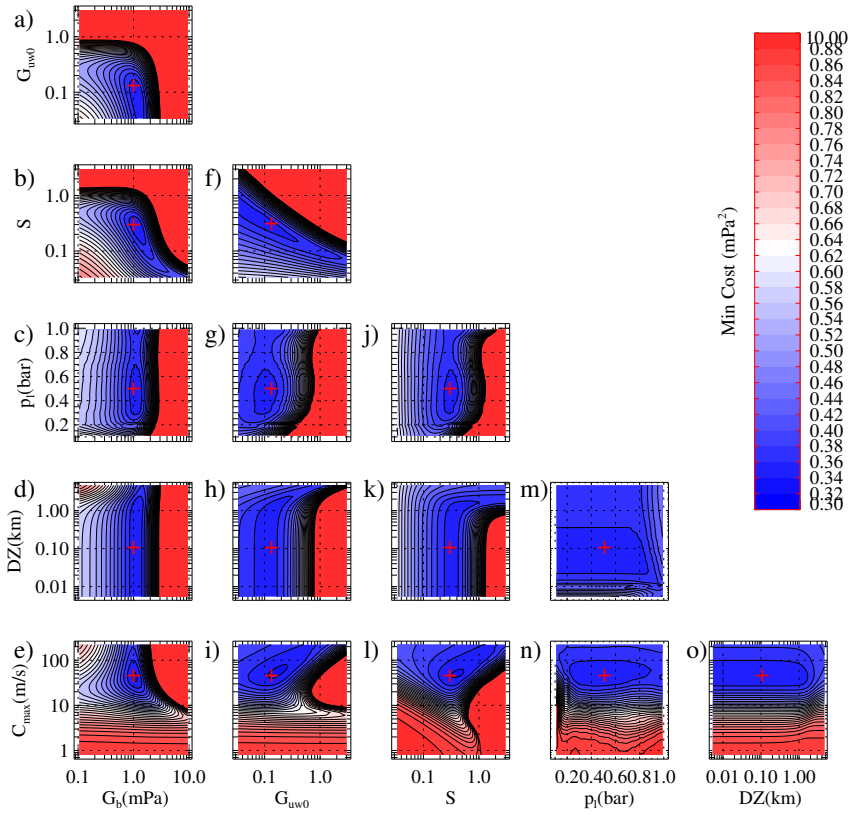

FIGURE 3 Cost function as a function of 2 parameters when the other 4 parameter are fixed at the values that yield the minimum shown in Fig. 2.

vary much around the minima at ($G_{uw0} = 0.15, G_b = 1$mPa) when G_{uw0} varies, but it varies significantly with the background G_b. Similarly, in the panels Figs. 2e)-2i)-2l)-2n)-2o), the changes in G_{uw0}, S, p_l and Δz do not affect the minima in the cost function. This is confirmed in the middle panels where these four parameters are varied separately, in all cases the minima in the cost do not vary much.

The fact that many different combinations of the six parameters ($G_{uw0}, S, \Delta z, p_l, C\text{max}, G_b$) can yield relatively small values of the cost function does not mean that the cost function does not vary when one of these parameters varies alone. It only means that when two parameters among these four are fixed, substantial minimization can be achieved by varying the 4 remaining ones. This suggests the presence of valleys in the cost function for some parameter combinations, in which the cost function varies slightly. In other words, the parameters are not all independent. This is illustrated in Fig. 3, which shows the minima in cost functions when two parameters vary and the other 4 parameter values are the optimal ones. The variations of the parameters around the minima are much more constrained, except for the heating source depth Δz, provided that $\Delta z < 1$km. This preliminary analysis indicates that, with this dataset, we cannot efficiently tune the 6 parameters of our scheme. To make a choice among the parameters, we next fix those that have a strong impact on the wave drag well above balloon flights, essentially where the waves are saturated and dynamical filtering dominates (two processes controlled by $C\text{max}$ and S). We therefore fix their values to their optimal ones (11), which are also quite near their operational value ($C\text{max} = 30\text{m/s} \rightarrow 40\text{m/s}$) and

$S = 0.6 \rightarrow 0.3$). We also choose the optimal value of Δz (Δz = 1km $\rightarrow$ 100m) but this is of little significance. In this paper, we will therefore focus on tuning the source amplitude and altitude only, G_{uwb}, G_{uw0}, and p_I.

4 | ESTIMATION TECHNIQUE

4.1 | EnKF/EnKS technique

To find optimal parameter values that, when applied to the GW scheme, permit the reproduction of the observed time series of momentum fluxes, we next use successive applications of an Ensemble Kalman filter combined with an expectation-maximization algorithm. This parameter estimation technique was presented in Tandeo et al. (2015).

The Ensemble Kalman filter technique is applied to each of the 23 Strateole flights independently, taking the logarithm of the daily averaged flux,

$$\boldsymbol{y}^{fT}(t_k) = \left[\log_{10}\left(F^{fE}\right), \log_{10}\left(F^{fW}\right)\right] = \mathcal{F}\left[\theta, Z(t_k)\right], \tag{12}$$

with similar expressions for the logarithm of the observation. We chose to work with logarithmic transformations because the logarithms of the fluxes are almost normally distributed (see Fig. 1c), as EnKF assumes normal distributions.

The GW scheme is used to calculate momentum fluxes and to conduct a daily forecast of the i^{th} member of an ensemble of N forecasts:

$$\boldsymbol{y}_i^f(t_k) = \mathcal{H}\{\boldsymbol{x}_i^f(t_k)\} = \mathcal{F}\left[\mathcal{G}\{\boldsymbol{x}_i^f(t_k)\}, \boldsymbol{Z}(t_k)\right]. \tag{13}$$

Thus, the gravity wave scheme acts as the observational operator through $\mathcal{F}$, projecting the forecasted state, $\boldsymbol{x}_i^f(t_k)$, to the observational space, i.e. momentum fluxes. The function $\mathcal{G}$ translates the Gaussian distribution of parameters $\boldsymbol{x}^f$, with typical values of order 1 into the parameters $\boldsymbol{\theta}$. This relation between the $\boldsymbol{x}$ and $\boldsymbol{\theta}$ is defined for each parameter as

$$\theta_{hi}(t_k) = \theta_{hL} + \frac{1. + \text{erf}\,(x_{hi}(t_k))}{2}\theta_{hU} \tag{14}$$

where θ_{hL} and θ_{hU} provide the lower and upper bound of the h^{th} parameter respectively. The estimate of $\mathbf{x}$ is then transformed to θ using (14).

Typically, the parameters at the initial time $\boldsymbol{x}_i^f(t_k)$ are initialized from a Gaussian distribution with mean $\boldsymbol{x}^b$ and covariance matrix $\boldsymbol{B}$. Thereafter, the parameters $\boldsymbol{x}_i^f$ are evolved each day according to a Gaussian random walk,

$$\boldsymbol{x}_i^f(t_k) = \boldsymbol{x}_i^a(t_{k-1}) + \boldsymbol{\eta}(t_k) \tag{15}$$

where $\boldsymbol{x}^a(t_{k-1})$ is the analysis at the previous time and $\boldsymbol{\eta}(t_k)$ is a Gaussian noise with zero mean and covariance matrix $\boldsymbol{Q}$, $\boldsymbol{\eta} = \mathcal{N}(0, \boldsymbol{Q})$. The EnKF also requests the specification of the observation error covariance matrix $\boldsymbol{R}$. We accumulate all the errors in observations and model them into $\boldsymbol{\epsilon}_i(t_k)$ as

$$\boldsymbol{y}^o(t_k) = \boldsymbol{y}_i^f(t_k) + \boldsymbol{\epsilon}_i(t_k), \tag{16}$$

where $\boldsymbol{\epsilon}_i(t_k) = \mathcal{N}(0, \boldsymbol{R})$.

At each time and from the ensemble we define sample mean unbiased estimates of the data and parameters

$$\overline{\boldsymbol{y}}^f(t_k) = \frac{1}{N}\sum_{i=1}^{N} \boldsymbol{y}_i^f(t_k), \quad \overline{\boldsymbol{x}}^f(t_k) = \frac{1}{N}\sum_{i=1}^{N} \boldsymbol{x}_i^f(t_k) \tag{17}$$

and the empirical covariance matrices,

$$\boldsymbol{P}^f_{\boldsymbol{rs}}(t_k) = \frac{1}{N-1}\sum_{i=1}^{N} \left\{\boldsymbol{r}_i^f(t_k) - \boldsymbol{r}^f(t_k)\right\}\left\{\boldsymbol{s}_i^f(t_k) - \boldsymbol{s}^f(t_k)\right\}^T, \tag{18}$$

where $\boldsymbol{r}$ and $\boldsymbol{s}$ stands for $\boldsymbol{x}$ or $\boldsymbol{y}$ and with similar expressions for the analysed and smoothed fields, in which case the f-superscript is replaced by a and s respectively. The analysis is then estimated recursively,

$$\boldsymbol{x}_i^a(t_k) = \boldsymbol{x}_i^f(t_k) + \boldsymbol{K}(t_k)\underbrace{\left(\boldsymbol{y}^o(t_k) - \boldsymbol{y}_i^f(t_k) + \boldsymbol{\epsilon}_i(t_k)\right)}_{\boldsymbol{d}_i(t_k)} \tag{19}$$

where $\boldsymbol{d}_i(t_k)$ is the innovation vector and $\boldsymbol{K}(t_k)$ is the Kalman gain,

$$\boldsymbol{K}(t_k) = \boldsymbol{P}^f_{\boldsymbol{xy}}(t_k)\left\{\boldsymbol{P}^f_{\boldsymbol{yy}}(t_k) + \boldsymbol{R}\right\}^{-1}. \tag{20}$$

Once the sequence in time is finished, we proceed backward to determine the smoothed estimates of state by applying a Kalman smoother,

$$\boldsymbol{x}_i^s(t_k) = \boldsymbol{x}_i^a(t_k) + \boldsymbol{K}^s(t_k)\left(\boldsymbol{x}_i^s(t_{k+1}) - \boldsymbol{x}_i^f(t_{k+1})\right), \tag{21}$$

with $\boldsymbol{x}_i^s(t_K) = \boldsymbol{x}^a(t_K)$, and where the Kalman smoother gain is,

$$\boldsymbol{K}^s(t_k) = \boldsymbol{P}^a_{\boldsymbol{xx}}(t_k)\left\{\boldsymbol{P}^f_{\boldsymbol{xx}}(t_{k+1})\right\}^{-1} \tag{22}$$

The application of this EnKF-EnKS technique requires the specification of the parameters at the initial time and its uncertainty, $\boldsymbol{x}^b$, $\boldsymbol{B}$, the model error $\boldsymbol{Q}$ which for the case of the GW scheme and the observation uncertainty of momentum fluxes, $\boldsymbol{R}$, are largely uncertain so that we propose to estimate them via the Expectation-Maximisation algorithm.

4.2 | Expectation-Maximisation algorithm

We apply an EM algorithm to estimate the statistical parameters $\boldsymbol{x}^b$, $\boldsymbol{B}$, $\boldsymbol{Q}$ and $\boldsymbol{R}$. The Expectation-Maximisation algorithm aims to iteratively maximize the likelihood function as a function of the statistical parameters. These statistical parameters are corrected during successive applications of the EM algorithm. At the first EM iteration, the statistical parameters are taken to be

$$\boldsymbol{x}^b = 0, \quad \boldsymbol{B} = 0.25 I_p, \quad \boldsymbol{Q} = 0.025 I_p \quad \boldsymbol{R} = 0.2 I_o; \tag{23}$$

where $\boldsymbol{I}$ is the identity matrix, which ranks o and p, the dimensions of the observation and parameter vector.

Given an initial set of statistical parameters $\boldsymbol{x}^b$, $\boldsymbol{B}$, $\boldsymbol{Q}$, and $\boldsymbol{R}(t_k)$, the Kalman smoother is used to an optimal state, $\boldsymbol{x}^s(t_k)$. In a given iteration, the realizations are given by the Kalman smoother $\boldsymbol{x}_i^s(t_k)$, and these are projected to the observational space by

$$\boldsymbol{y}_i^s(t_k) = \mathcal{H}\{\boldsymbol{x}_i^s(t_k)\}. \tag{24}$$

The maximum likelihood estimates of the statistical parameters $\boldsymbol{x}^b$, $\boldsymbol{B}$, $\boldsymbol{Q}$, and $\boldsymbol{R}(t_k)$ are then evaluated using $\boldsymbol{x}_i^s(t_k)$ and $\boldsymbol{y}_i^s(t_k)$,

$$\hat{\boldsymbol{x}}_b = \boldsymbol{x}^s(t_1),\ \hat{\boldsymbol{B}} = \boldsymbol{P}_{\boldsymbol{xx}}^s(t_1) \tag{25}$$

$$\hat{\boldsymbol{Q}} = \frac{1}{K-1}\frac{1}{N}\sum_{k=2}^{K}\sum_{i=1}^{N}\left\{\boldsymbol{x}_i^s(t_k) - \boldsymbol{x}_i^s(t_{k-1})\right\}\left\{\boldsymbol{x}_i^s(t_k) - \boldsymbol{x}_i^s(t_{k-1})\right\}^T \tag{26}$$

$$\hat{\boldsymbol{R}} = \frac{1}{K}\frac{1}{N}\sum_{k=1}^{K}\sum_{i=1}^{N}\left\{\boldsymbol{y}_i^s(t_k) - \boldsymbol{y}^o(t_k)\right\}\left\{\boldsymbol{y}_i^s(t_k) - \boldsymbol{y}^o(t_k)\right\}^T \tag{27}$$

A derivation of these expressions may be found in Tandeo et al. (2015) and Pulido et al. (2018). Once a new set of statistical parameters is estimated, we then return to (23) for a new iteration until a minimum in the innovation negative log-likelihood estimate,

$$\mathcal{L} = \frac{1}{K-1}\frac{1}{N}\sum_{k=1}^{K}\sum_{i=1}^{N}\left\{\boldsymbol{d}_i(t_k)^T\left(\boldsymbol{P}_{yy}^f(t_k) + \boldsymbol{R}\right)\boldsymbol{d}_i(t_k) + \frac{1}{2}\log\left(\det[\boldsymbol{P}_{yy}^f(t_k) + \boldsymbol{R}]\right)\right\} \tag{28}$$

is reached.

5 | RESULTS

5.1 | Estimation of launching pressure p_l and background wave amplitude G_{uwb}

In order to evaluate the EnKF/EM technique using real data, we estimate the launching pressure and background amplitude using the balloon data. The launching pressure is limited to altitudes below balloon flight and large bounds are taken for the background amplitude, namely

$$\theta_4 = p_l, \text{ with } (\theta_{4L}, \theta_{4U}) = (0.1, 1) \text{ bar, and } \theta_6 = G_{uwb} \text{ with } (\theta_{6L}, \theta_{6U}) = (0, 10) \text{ mPa,} \tag{29}$$

the other parameter values are given in (11). The reason for selecting two parameters is that the daily data of observed and forecast MF are not decorrelated from one day to the next. The typical decorrelation time is around 5 day (not shown) reducing the number of degree of freedom of the observations to only a few hundred. In this case, tuning more than two parameters is questionable, we therefore proceed by pairs and estimate the coevolution between G_{uwb} and G_{uw0} for fixed p_l thereafter. Note for completeness that we conducted preliminary experiments using the EnKF/EM

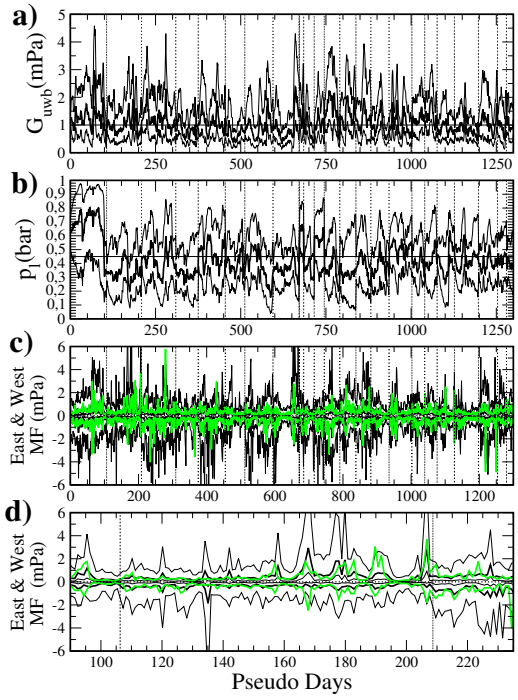

FIGURE 4 EnKF estimates of the two parameters (G_{uwb} and x_l). a) $G^f_{uwb}(t_k)$ (thick solid) and $\left(G^f_{uwb} \pm 2\sigma(G^f_{uwb})\right)(t_k)$ (thin solid), where the variance σ is evaluated from the ensemble forecast. b) Same as a) but for p_l. c) Same as a) but for East and West MFs resulting from EnKF estimates (black), and observed values are in green. (d) East and West MF zoom view for a range of days.

algorithm for the G_{uwb}, G_{uw0}, and p_l altogether, and the results are not much different.

Figure 4 shows the variations over balloon flight and over time of the two parameters G_{uwb} and p_l after 4 EM iterations and using an ensemble size of $N = 100$. The choice of stopping the analysis at 4 EM iterations for each balloon is motivated by the fact that for most balloons the innovation log-likelihood (28) has almost reached a minimum at that iteration. It emphasizes that in our case the algorithm converges quite fast. The top and middle panels of Fig. 4 show the temporal variations of the ensemble average of G^f_b and X^f_l, each bounded by a 2-standard deviation. Both parameters oscillate around mean values that are almost equal to the optimal values identified at the extrema of the cost function. Hence, the EnKF/EM permits estimating the optimal values of these two parameters. Substantial variations around the entire dataset mean are found for both the averaged forecasts and the standard deviation, with a launching altitude easily covering the entire troposphere whereas the background amplitude is constrained between 0.3mPa and 3mPa typically. Fig. 4 (c) shows the temporal evolution of the forecast $y^f(t_k)$, bounded by two standard deviations. It successfully captures the variations in the measured data, as almost all measured momentum fluxes fall within the $\pm 2\sigma$ curves. The zoom view in panel (d) reveals that the forecast also has peak values whose amplitudes compare well to the measured peak values, albeit with some minor timing discrepancies.

To confirm more quantitatively that the ensemble captures well the spread in observations, we compute probabil-

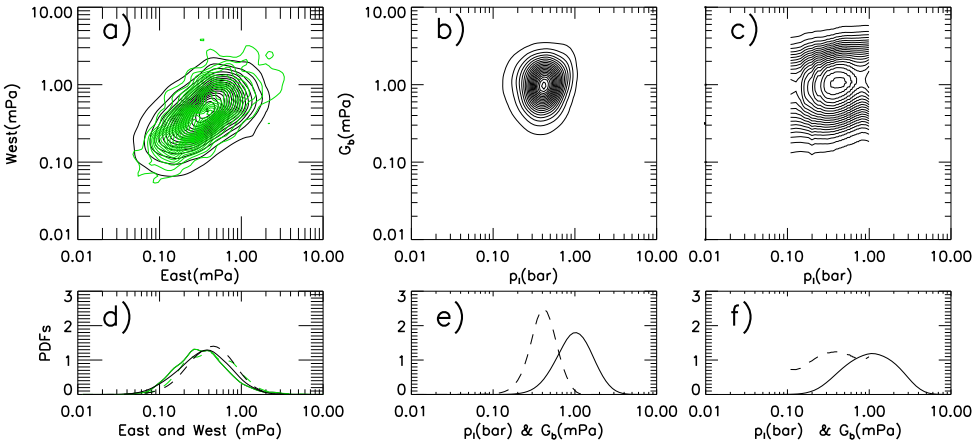

FIGURE 5 PDFs of momentum fluxes and parameters. (a) PDFs of the East and West MFs from the ensemble forecast (black) and observed (green) and (d) the marginal distributions. b), c), e), and f) PDFs of launching altitude and launched background flux(x_l and G_b) : b) and e) from the EnKF forecasts at the last EM iteration; c) and f) from Bayesian inference using model error PDFs build with Gaussian Kernels.

ity distributions using radial basis functions or Gaussian kernels (Wand and Jones, 1994), and estimate the probability density distribution of the daily fluxes as

$$P^o(\tilde{\mathbf{y}}) = \frac{1}{2\pi K \delta y^2} \sum_{k=1}^{K} \exp\left(-(\mathbf{y}-\mathbf{y}^o(t_k))(\mathbf{y}-\mathbf{y}^o(t_k))^T/2\delta y^2\right) \tag{30}$$

$$P^{\mathrm{EnKf}}(\tilde{y}) = \frac{5}{2\pi K N \delta y^2} \sum_{i=1}^{N} \sum_{k=3,5}^{K} \exp\left(-(\tilde{\mathbf{y}}-\tilde{\mathbf{y}}_i^f(t_k))(\tilde{\mathbf{y}}-\tilde{\mathbf{y}}_i^f(t_k))^T/2\delta y^2\right) \tag{31}$$

the bandwidth δy = 0.1mPa typically. Note that formula (30) can be applied to individual forecasts changing $\boldsymbol{y}^o$ by $\boldsymbol{y}^f$ in (30) and that (31) can be applied with similar resolution to the $\log_{10}$ of p_l (in bar) and G_b (in mPa) yielding a probability distribution of parameters $P^{\mathrm{EnKf}}(\boldsymbol{\theta})$.

Figure 5a) shows the PDF of the MFs from the observations in green compared to the PDF from the ensemble in black. The ensemble resulting from EnKF/EM reproduces rather well the observation statistics of the parameters. This is essential if one wishes to provide quantitative estimates of the parameter variability. Particularly, the observation PDF is anisotropic and shows substantial correlation between the Eastward and Westward MFs, both features are rather well reproduced by the ensemble. To identify the cause, we have conducted experiments with the EnKF/EM varying only one of the parameters, G_{uwb} or p_l and found that the variations in the former rather than the variations in the launching altitude are the ones that produce the correlation.

The PDF of the parameter from the EnKF are shown in Figs. 5b) and 5e). As expected from the curves in Figs. 4 one sees that the launching altitude is centered in the mid troposphere but can be almost everywhere in the troposphere (e.g. $p_l > 0.1$) and that the background amplitude can have any values between 0.3mPa< G_b < 3mPa. This is quite high variability, that may call for further refinement in the parameterization, or on the calculation of the MFs from the data, or on the collocation of the ERA5 fields, but we leave this to future work. For the purposes of this work,

however, we are primarily concerned with assessing the utility of these PDFs for uncertainty quantification of the parameters.

5.2 | Direct Bayes inversion using kernel density estimates of the model error PDF

To evaluate the inferred PDFs, we can find the posterior distribution using a direct analysis by varying uniformly p_l to cover the troposphere ($1 > p_l > 0.1$) and logarithmically G_b between 0.01mPa and 10mPa and evaluating the GW scheme for these parameters; in this way, $50 \times 50 = 2500$ simulations are conducted for the entire Strateole 2 data. In other words, we assume a uniform prior distribution in the parameters. Next, we build PDFs of the model error ϵ using Gaussian Kernels and for the best model outputs the parameter value θ_M:

$$P_\epsilon(\epsilon) = \frac{1}{2\pi K \delta y^2} \sum_{k=1}^{K} e^{\left(-(\epsilon-(\mathbf{y}^o(t_k)-\mathbf{y}^f(t_k,\theta_M)))(\epsilon-(\mathbf{y}^o(t_k)-\mathbf{y}^f(t_k,\theta_M)))^T/(\delta y)^2\right)}. \tag{32}$$

From the density probability of the model errors we then build the likelihood of y when θ is fixed as,

$$P(y|\theta) = P_\epsilon\left(y - y^f(\theta)\right). \tag{33}$$

The posterior distribution for $\boldsymbol{\theta}$ given the observations $\{y^o(t_k)\}_{k=1}^K$ can be written as,

$$P(\theta|y^o) \propto \prod_{k=1}^{K} P_\epsilon\left(y^o(t_k) - y^f(t_k,\theta)\right) p(\theta), \tag{34}$$

where $p(\theta)$ is the prior distribution, which is assumed uniform. This expression assumes that the errors are uncorrelated from one day to the other. Note that this last assumption is not well verified, so we checked the results from (34) taking one daily data every 3 days, or more to enforce independence, which does not change the results much.

This direct estimation of the posterior distribution in Figs. 4c) and 4f) has some similarities with the EnKF parameter PDF in Figs. 4b) and 4e) being broader in both the launching altitude p_l and background flux amplitude G_b. To interpret the difference, we have evaluated variants of the EnKF and the "direct" PDFs estimate by analysing slower variations. Such tests are motivated by the fact that daily data of fluxes (observed and forecast) and parameters from the EnKF have substantial lag correlations and become decorrelated at lags of about 5 days. We therefore recompute PDFs using data averaged over 5 days after applying cosine tapers to enforce decorrelation. Applying such low-pass filters, we found that the statistics of the EnKF parameters are not much different, whereas the Bayesian estimate of the posterior distribution narrows for the amplitude G_b getting closer to the EnKF estimate but it remains unchanged for the launching altitude p_l (not shown). Concerning G_b, it means that EnKF estimates are more robust than direct estimates since EnKF takes into account the lag correlations in the data. For the launching pressure p_l on the other hand, the direct Bayesian estimate always shows a broad distribution over the entire troposphere, with a relative minimum in the middle, whereas the EnKF proposes a normal distribution around this minimum. This naturally follows that EnKF theory explicitly assumes Gaussian statistics: the launching GWs altitude being more uniformly distributed over the troposphere, so that its statistics is not well captured by the EnKF.

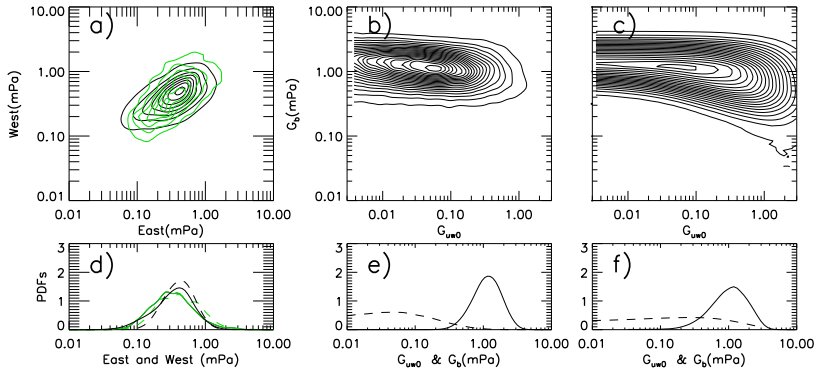

FIGURE 6 PDFs of Momentum fluxes and parameters. a) East and West MFs from the ensemble forecast (black) and observed (green) and (d) the marginal distributions. b), c), e), and f) PDFs of convective wave parameter and background flux(G_{uw0} and G_b) : b) and e) from the ENKF forecasts at the last EM iteration; c) and f) from Bayesian inference using model error PDFs build with Gaussian Kernels.

5.3 | Results for G_{uw0} and G_b

We conducted a similar experiment for the amplitude parameters G_{uw0} and G_b. The resulting PDFs are shown in Fig. 6. Again the east and west fluxes from the EnKF PDFs match quite well the observed PDFs (Figs. 6a) and 6d). The PDFs of the parameters G_{uw0} and G_b in Figs. 6b) and 6e) indicate that G_{uw0} can take almost any value below $G_{uw0} < 0.3$ typically, provided that 0.3mPa $< G_b <$ 3mPa. This clearly highlights a difficulty in tuning the parameter that controls the convective source. One potential issue is that the EnKF may struggle to handle precipitation data, for instance, by continuing to adjust the forecast for the corresponding parameter even when no precipitation is present. We tested this potential issue conducting the EnKF sequentially but only assimilating data when there is precipitation. The result show little difference. The problem seems to be related to the sources, which are a function of the square of the precipitation (P^2, see Eq. 4). This is confirmed in Figs.6c) and 6d) which show that the direct Bayesian estimate predicts about the same statistical behavior for $G_{\mathbf{uw0}}$. Both analyses demonstrate that G_{uw0} can take any values less than 1 provided that G_b stays within its bounds.

6 | SUMMARY AND DISCUSSION

6.0.1 | Summary

In this article, we use an offline filtering technique to estimate the physical parameters of a subgrid scale non-orographic gravity wave parameterization. The technique was originally presented and tested with synthetic observations in Tandeo et al. (2015). Its main advantages are that it is designed to solve a nonlinear space-state system using the Ensemble Kalman Filter, which is readily adapted to a gravity wave parameterization, and that it estimates the unknown error covariance matrices. More precisely, this technique assumes the hypothesis that the state, the observation, and the parameter equations have additive Gaussian noises, whose statistics are estimated via an iterative maximum likelihood method using successive applications of the EnKF.

Here, for the first time, the method is applied to real observations of the gravity wave momentum fluxes. These fluxes were deduced from constant-altitude balloon flights conducted during the Strateole 2 campaign. Unlike the

synthetic observation experiments in Tandeo et al. (2015), the optimal parameter values are not known a priori in experiments with real data. Therefore, we validate the results from the filtering technique against a direct optimization conducted using a brute-force technique that systematically varies the physical parameters of the GW scheme.

After identifying that for this dataset and parameterization, the parameter variations are not independent, we conclude that the optimization can only be conducted for 2 to 3 parameters keeping the others fixed, for instance taking their operational values. This choice regarding the number of estimable parameters is further supported by a lag correlation analysis of the data. In fact, the daily mean of winds from ERA5 as well as daily mean of the measured momentum fluxes present substantial lag correlations for up to 3-5 days. This reduces the number of degrees of freedom of our dataset substantially from 1294 days of measurements to around 200. For these reasons, we chose to fix the parameters that play a central role in the dynamical filtering of the gravity waves above the balloon altitude at their reference value, (C_{Max} and S). We also fixed the source depth Δz, as we found the parameterization is largely insensitivity to its value provided that $\Delta Z < 1$km. We therefore focused our analysis on the remaining parameters related to the tropospheric sources, the launching altitude, z_l, the convective gravity wave amplitude G_{uw0} and the background gravity wave activity G_b. A key result of this paper is that, for these selected parameters, the technique permits to find their optimal values and to provide estimates of their associated uncertainties. To reach this conclusion, the EnKf/EM method is compared to a direct method. In the direct method, PDFs of the error are built from a large ensemble of parameter realisations. In this case, the optimal values are found by minimizing the cost function via a direct technique, and the uncertainties are inferred by calculating the posterior distribution using Bayesian methods with Gaussian Kernels.

Several findings concerning the parameterization itself are worth noting. A first result is that the preferential location of the gravity wave source is in the mid troposphere. The uncertainty is nevertheless large, both EnKf and direct Bayesian inference show that gravity waves can be launched from almost any other place in the troposphere below the balloon flight level. The EnKF/EM method produces a normal distribution of the launching altitude around the optimal value, whereas the direct Bayesian inference gives a distribution closer to the uniform one. Regarding the background amplitude, the analysis identifies an optimum amplitude of $G_b = 1.25$mPa, but shows that almost any value within the range $0.3 < G_b < 3$mPa is similarly plausible. The experiment, in which the convective wave amplitude and the background parameters G_{uw0} and G_b are estimated, conveys a similar message to the last one. For the convective wave amplitude it shows that it can have almost any value provided that $G_{uw0} < 1$ the background amplitude is within the right bounds. The operational situation of a null background amplitude is rejected, making the introduction of a background source mandatory. That said, we note nevertheless that the LMDz operational value $G_{uw0} = 0.25$ falls within the plausible bounds compared to the observational bounds. This suggests that a background of wave activity can potentially be introduced operationally without being overly detrimental to the model's performance.

6.1 | Discussion

We conducted a large number of sensitivity experiments, changing the ensemble size N, not taking the logarithm of the fluxes, considering net and amplitude fluxes rather than eastward and westward. The results show little qualitative difference from the above conclusions. We have also tested the other parameters and found broad ranges of plausible values. For instance, C_{max} can easily be larger than 50m/s, yielding to faster gravity waves than expected. This call to collect data at higher altitudes to include more substantial dynamical filtering in the analysis.

Concerning the level of uncertainties we found, it seems to be quite large compared to when parameters are changed in online experiments in Garfinkel et al. (2022) and Mansfield and Sheshadri (2022). For instance, Mansfield and Sheshadri (2022) used another gravity wave parameterization and applied the EnKF comparing online GCMs run

to observed QBO statistics, and found uncertainties on the source term much narrower than ours on the background amplitude for instance. This difference is noteworthy. Preliminary tests show that the offline uncertainties we find could be substantially narrowed if we used longer time averaged momentum fluxes. This would be more consistent in terms of the QBO dynamics which forcing integrates the MFs over long time scales. A first problem is that in this case the number of in-situ data we have in hand becomes much more limited. A second one is that when averaging data over long time intervals, we lose the chance to relate the variations in gravity wave amplitude to their sources at a given time and place. A possibility to be explored in the future is to improve the relation with the sources. This being said, we have started to low pass the outcomes from the EnKF, using bow-car windows up to 10 day. This decrease the noise and the uncertainty, but not to the level found in Mansfield and Sheshadri (2022). This may show that models are over-sensitive to parameter changes, but it may also show that the level of noise and error introduced by the extraction of momentum fluxes using balloon data, the collocation of the synoptic winds, and in the ERA5 fields themselves stay too large to narrow the uncertainties about our optimal parameter estimate.

In future work, we plan to test if our results can help improve the parameterization in operation. For example, we will examine whether the introduction of a background wave source is beneficial to the IPSLCM6 model. Preliminary tests show that it cannot fully replace the convective source, i.e that we should keep $G_{uw0} \neq 0$ at least when a constant value for the background G_b is considered: the high degree of intermittency resulting from the convective sources seems to help the model to simulate a QBO. Also, and on top of uncertainties due to observations and model errors, a fundamental reason why it is not possible to tune fixed parameter values is that none of them is supposed to be a fixed universal constant. Gravity wave sources, propagation, and breaking depend on local flow conditions and forcing structures that cannot be predicted at each time and place. There are fundamental reasons for that, one is that the forced mesoscale flow radiating gravity waves is fully turbulent and not predictable. Another one is that the gravity wave field has time and space scales quite near the resolved scales, it is far from a statistical equilibrium that would allow to derive fixed scaling coefficients. These fundamental reasons have justified the development of stochastic parameterizations (Berner et al., 2017) in the past. In this framework, our analysis could be used to make the background stochastic, using statistics of our forecasts of G_b for instance. To try refining our parameterization further we have also started to analyse potential relations between the changes in the background amplitude G_b with the flow profiles and precipitations to see if it varies according to identifiable causes.We also leave this for further analysis, but preliminary results indicate that larger amplitude fluxes should be imposed in the presence of larger amplitude shears in the mid troposphere and that the P^2 dependence in the convective fluxes in (4) could be reconsidered.

Acknowledgements

We thank Iman Toghraei and Raj Rani for their constructive feedbacks.

Data availability statement

The datasets, GW parameterization and EnKF/EM algorithm used to support the findings of this study are available from the corresponding author webpage(`https://web.lmd.jussieu.fr/~flott/publi_data.html`)

Conflict of interest statement

The authors declare no conflict of interest.

references

Achatz, U., Kim, Y.-H. and Voelker, G. S. (2023) Multi-scale dynamics of the interaction between waves and mean flows: From nonlinear WKB theory to gravity-wave parameterizations in weather and climate models. *Journal of Mathematical Physics*, **64**, 111101. URL: https://doi.org/10.1063/5.0165180.

Alexander, M. J., Geller, M., McLandress, C., Polavarapu, S., Preusse, P., Sassi, F., Sato, K., Eckermann, S., Ern, M., Hertzog, A., Kawatani, Y., Pulido, M., Shaw, T. A., Sigmond, M., Vincent, R. and Watanabe, S. (2010) Recent developments in gravity-wave effects in climate models and the global distribution of gravity-wave momentum flux from observations and models. *Q. J. R. Meteorol. Soc.*, **136**, 1103–1124.

Andrews, F. G., Holton, J. and Leovy, C. (1987) *Middle Atmosphere Dynamics*. Academic Press.

Annan, J., Hargreaves, J., Edwards, N. and Marsh, R. (2005) Parameter estimation in an intermediate complexity earth system model using an ensemble kalman filter. *Ocean Modelling*, **8**, 135–154.

Anstey, J. A., Scinocca, J. F. and Keller, M. (2016) Simulating the qbo in an atmospheric general circulation model: Sensitivity to resolved and parameterized forcing. *Journal of the Atmospheric Sciences*, **73**, 1649 – 1665.

Baldwin, M. P., Gray, L. J., Dunkerton, T. J., Hamilton, K., Haynes, P. H., Randel, W. J., Holton, J. R., Alexander, M. J., Hirota, I., Horinouchi, T., Jones, D. B. A., Kinnersley, J. S., Marquardt, C., Sato, K. and Takahashi, M. (2001) The quasi-biennial oscillation. *Rev. Geophys.*, **39**, 179–229.

Beres, J. H., Garcia, R. R., Boville, B. A. and Sassi, F. (2005) Implementation of a gravity wave source spectrum parameterization dependent on the properties of convection in the whole atmosphere community climate model (waccm). *Journal of Geophysical Research: Atmospheres*, **110**.

Berner, J., Achatz, U., Batté, L., Bengtsson, L., de La Cámara, A. Christensen, H. M., Colangeli, M., Coleman, D. R. B., Crommelin, D., Dolaptchiev, S. I., Franzke, C. L. E., Friederichs, P., Imkeller, P., Järvinen, H., Juricke, S., Kitsios, V., Lott, F., Lucarini, V., Mahajaajaajan, S., Palmer, T. N., Penland, C., Sakradzijaja, M., Von Storch, J.-S., Weisheimer, A., Weniger, M., Williams, P. D. and Yano, J.-I. (2017) Stochastic parameterization toward a new view of weather and climate models. *Bull. Amer. Meteor. Soc.*, **98**, 565–587.

Boucher, O., Servonnat, J., Albright, A. L., Aumont, O., Balkanski, Y., Bastrikov, V., Bekki, S., Bonnet, R., Bony, S., Bopp, L., Braconnot, P., Brockmann, P., Cadule, P., Caubel, A., Cheruy, F., Codron, F., Cozic, A., Cugnet, D., D'Andrea, F., Davini, P., de Lavergne, C., Denvil, S., Deshayes, J., Devilliers, M., Ducharne, A., Dufresne, J.-L., Dupont, E., Éthé, C., Fairhead, L., Falletti, L., Flavoni, S., Foujols, M.-A., Gardoll, S., Gastineau, G., Ghattas, J., Grandpeix, J.-Y., Guenet, B., Guez, Lionel, E., Guilyardi, E., Guimberteau, M., Hauglustaine, D., Hourdin, F., Idelkadi, A., Joussaume, S., Kageyama, M., Khodri, M., Krinner, G., Lebas, N., Levavasseur, G., Lévy, C., Li, L., Lott, F., Lurton, T., Luyssaert, S., Madec, G., Madeleine, J.-B., Maignan, F., Marchand, M., Marti, O., Mellul, L., Meurdesoif, Y., Mignot, J., Musat, I., Ottlé, C., Peylin, P., Planton, Y., Polcher, J., Rio, C., Rochetin, N., Rousset, C., Sepulchre, P., Sima, A., Swingedouw, D., Thiéblemont, R., Traore, A. K., Vancoppenolle, M., Vial, J., Vialard, J., Viovy, N. and Vuichard, N. (2020) Presentation and evaluation of the ipsl-cm6a-lr climate model. *Journal of Advances in Modeling Earth Systems*, **12**, e2019MS002010.

Bushell, A. C., Butchart, N., Derbyshire, S. H., Jackson, D. R., Shutts, G. J., Vosper, S. B. and Webster, S. (2015) Parameterized gravity wave momentum fluxes from sources related to convection and large-scale precipitation processes in a global atmosphere model. *Journal of the Atmospheric Sciences*, **72**, 4349–4371.

Carrassi, A., Bocquet, M., Bertino, L. and Evensen, G. (2018) Data assimilation in the geosciences: An overview of methods, issues, and perspectives. *Wiley Interdisciplinary Reviews: Climate Change*, **9**, e535.

Christiansen, B., Yang, S. and Madsen, M. S. (2016) Do strong warm enso events control the phase of the stratospheric qbo? *Geophysical Research Letters*, **43**, 10,489–10,495.

Cleary, E., Garbuno-Inigo, A., Lan, S., Schneider, T. and Stuart, A. M. (2021) Calibrate, emulate, sample. *Journal of Computational Physics*, **424**, 109716.

Corcos, M., Hertzog, A., Plougonven, R. and Podglajen, A. (2021) Observation of gravity waves at the tropical tropopause using superpressure balloons. *Journal of Geophysical Research: Atmospheres*, **126**, e2021JD035165.

Couvreux, F., Hourdin, F., Williamson, D., Roehrig, R., Volodina, V., Villefranque, N., Rio, C., Audouin, O., Salter, J., Bazile, E., Brient, F., Favot, F., Honnert, R., Lefebvre, M.-P., Madeleine, J.-B., Rodier, Q. and Xu, W. (2021) Process-based climate model development harnessing machine learning: I. a calibration tool for parameterization improvement. *Journal of Advances in Modeling Earth Systems*, **13**, e2020MS002217.

Dempster, A. P., Laird, N. M. and Rubin, D. B. (1977) Maximum likelihood from incomplete data via the em algorithm. *Journal of the Royal Statistical Society: Series B (Methodological)*, **39**, 1–22.

Dunbar, O. R. A., Garbuno-Inigo, A., Schneider, T. and Stuart, A. M. (2021) Calibration and uncertainty quantification of convective parameters in an idealized gcm. *Journal of Advances in Modeling Earth Systems*, **13**, e2020MS002454.

Ern, M., Ploeger, F., Preusse, P., Gille, J., Gray, L. J., Kalisch, S., Mlynczak, M. G., Russell, J. M. and Riese, M. (2014) Interaction of gravity waves with the QBO: A satellite perspective. *Journal of Geophysical Research: Atmospheres*, **119**, 2329 – 2355.

Evensen, G. (2009) *Data Assimilation: The Ensemble Kalman Filter*. Spinger-Verlag.

Garfinkel, C. I., Gerber, E. P., Shamir, O., Rao, J., Jucker, M., White, I. and Paldor, N. (2022) A QBO cookbook: Sensitivity of the quasi-biennial oscillation to resolution, resolved waves, and parameterized gravity waves. *Journal of Advances in Modeling Earth Systems*, **14**, e2021MS002568.

Geller, M. A., Alexander, M. J., Love, P. T., Bacmeister, J., Ern, M., Hertzog, A., Manzini, E., Preusse, P., Sato, K., Scaife, A. A. and Zhou, T. (2013) A comparison between gravity wave momentum fluxes in observations and climate models. *J. Atmos. Sci.*, **26**.

Geller, M. A., Zhou, T., Shindell, D., Ruedy, R., Aleinov, I., Nazarenko, L., Tausnev, N. L., Kelley, M., Sun, S., Cheng, Y., Field, R. D. and Faluvegi, G. (2016) Modeling the qbo—improvements resulting from higher-model vertical resolution. *Journal of Advances in Modeling Earth Systems*, **8**, 1092–1105.

Giorgetta, M. A., Manzini, E., Roeckner, E., Esch, M. and Bengtsson, L. (2006) Climatology and forcing of the quasi-biennial oscillation in the maecham5 model. *Journal of Climate*, **19**, 3882 – 3901.

Green, B., Sheshadri, A., Alexander, M., Bramberger, M. and Lott, F. (2023) Gravity wave momentum fluxes estimated from project loon balloon data. *Journal of Geophysical Research: Atmospheres*, **Submitted**.

Haase, J. S., Alexander, M. J., Hertzog, A., Kalnajs, L. E., Deshler, T., Davis, S. M., Plougonven, R., Cocquerez, P. and Venel, S. (2018) Around the world in 84 days [Dataset]. *Eos*, **99**.

Hertzog, A., Alexander, M. J. and Plougonven, R. (2012) On the Intermittency of Gravity Wave Momentum Flux in the Stratosphere. *Journal of the Atmospheric Sciences*, 3433–3448.

Jewtoukoff, V., Plougonven, R. and Hertzog, A. (2013) Gravity waves generated by deep tropical convection: Estimates from balloon observations and mesoscale simulations. *Journal of Geophysical Research: Atmospheres*, **118**, 9690–9707.

Lott, F. and Guez, L. (2013) A stochastic parameterization of the gravity waves due to convection and its impact on the equatorial stratosphere. *J. Geophys. Res.*, **118**, 8897–8909.

Lott, F. and Miller, M. J. (1997) A new subgrid-scale orographic drag parametrization: Its formulation and testing. *Quarterly Journal of the Royal Meteorological Society*, **123**, 101–127. URL: `http://www.ingentaselect.com/rpsv/cgi-bin/cgi?ini=xref{\&}body=linker{\&}reqdoi=10.1256/smsqj.53703`.

Lott, F., Rani, R., Podglajen, A., Codron, F., Guez, L., Hertzog, A. and Plougonven, R. (2023) Direct comparison between a non-orographic gravity wave drag scheme and constant level balloons. *Journal of Geophysical Research: Atmospheres*, **128**, e2022JD037585.

Mansfield, L. A. and Sheshadri, A. (2022) Calibration and uncertainty quantification of a gravity wave parameterization: A case study of the quasi-biennial oscillation in an intermediate complexity climate model. *Journal of Advances in Modeling Earth Systems*, **14**, e2022MS003245.

— (2024) Uncertainty quantification of a machine learning subgrid-scale parameterization for atmospheric gravity waves. *Journal of Advances in Modeling Earth Systems*, **16**, e2024MS004292.

Orr, A., Bechtold, P., Scinocca, J., Ern, M. and Janiskova, M. (2010) Improved middle atmosphere climate and forecasts in the ecmwf model through a nonorographic gravity wave drag parameterization. *Journal of Climate*, **23**, 5905 – 5926.

Pulido, M., Tandeo, P., Bocquet, M., Carrassi, A. and Lucini, M. (2018) Stochastic parameterization identification using ensemble kalman filtering combined with maximum likelihood methods. *Tellus A: Dynamic Meteorology and Oceanography*, **70**, 1–17.

Ruiz, J. and Pulido, M. (2015) Parameter estimation using ensemble-based data assimilation in the presence of model error. *Monthly Weather Review*, **143**, 1568–1582.

Ruiz, J. J., Pulido, M. and Miyoshi, T. (2013) Estimating model parameters with ensemble-based data assimilation: A review. *Journal of the Meteorological Society of Japan*, **91**, 79–99.

Scinocca, J. F. (2003) An accurate spectral nonorographic gravity wave drag parameterization for general circulation models. *Journal of the Atmospheric Sciences*, **60**, 667 – 682.

Serva, F., Cagnazzo, C., Riccio, A. and Manzini, E. (2018) Impact of a stochastic nonorographic gravity wave parameterization on the stratospheric dynamics of a general circulation model. *Journal of Advances in Modeling Earth Systems*, **10**, 2147–2162.

Song, I.-S. and Chun, H.-Y. (2005) Momentum flux spectrum of convectively forced internal gravity waves and its application to gravity wave drag parameterization. part i: Theory. *J. Atmos. Sci.*, **62**, 107–124.

Tandeo, P., Ailliot, P., Bocquet, M., Carrassi, A., Miyoshi, T., Pulido, M. and Zhen, Y. (2020) A review of innovation-based methods to jointly estimate model and observation error covariance matrices in ensemble data assimilation. *Monthly Weather Review*, **148**, 3973–3994.

Tandeo, P., Pulido, M. and Lott, F. (2015) Offline parameter estimation using enkf and maximum likelihood error covariance estimates: Application to a subgrid-scale orography parametrization. *Quarterly Journal of the Royal Meteorological Society*, **141**, 383–395.

Voelker, G. S., Bölöni, G., Kim, Y.-H., Zängl, G. and Achatz, U. (2024) Ms-gwam: A three-dimensional transient gravity wave parametrization for atmospheric models. *Journal of the Atmospheric Sciences*, **81**, 1181 – 1200.

Wand, M. P. and Jones, M. C. (1994) *Kernel smoothing*. CRC press.