RMetS
Royal Meteorological Society

Offline parameter estimation using EnKF and maximum likelihood error covariance estimates: Application to a subgrid-scale orography parametrization

P. Tandeo,[a,b]* M. Pulido[a] and F. Lott[c]

[a]*Department of Physics, Universidad Nacional del Nordeste, Corrientes, Argentina*
[b]*Lab-STICC –Pôle CID, Telecom Bretagne, Brest, France*
[c]*Laboratoire de Meteorologie Dynamique, Ecole Normale Superieure, Paris, France*

*Correspondence to: P. Tandeo, Telecom Bretagne, 29280 Plouzané, France.
E-mail: pierre.tandeo@telecom-bretagne.eu

Recent work has shown that the parameters controlling parametrizations of the physical processes in climate models can be estimated from observations using filtering techniques. In this article, we propose an offline parameter estimation approach, without estimating the state of the climate model. It is based on the Ensemble Kalman Filter (EnKF) and an iterative estimation of the error covariance matrices and of the background state using a maximum likelihood algorithm. The technique is implemented in a subgrid-scale orography (SSO) parametrization scheme which works in a single vertical column. First, the parameter estimation technique is evaluated using twin experiments. Then, the technique is used with synthetic observations to estimate how the parameters of the SSO scheme should change when the resolution of the input orography dataset of a general circulation model is increased. Our analysis reveals that, when the resolution of the orography dataset increases, the scheme should take into account the dynamical sheltering that can occur at low levels between mountain peaks located within the same gridbox area.

Key Words: offline parameter estimation; EnKF; EM algorithm; subgrid-scale orography parametrization

Received 21 June 2013; Revised 4 February 2014; Accepted 27 February 2014; Published online in Wiley Online Library 15 April 2014

1. Introduction

Numerical models, including atmospheric/oceanic general circulation models (GCMs) and current earth system models, contain several physical parametrizations with a large number of parameters. Climate predictions using these numerical models are sensitive to the large set of parameters that are present in the physical parametrizations (cf. Stainforth *et al.*, 2005). Most of these unknown physical parameters can not be determined directly from observations and are generally manually tuned. This subjective approach is excessively time demanding and does not give optimal results. Moreover, if the horizontal resolution of the model or of an input dataset is increased or a parametrization scheme is changed, the physical parameters need to be re-evaluated. To address these issues, several authors (e.g. Jackson *et al.*, 2004; Severijns and Hazeleger, 2005) propose estimating the physical parameters objectively, defining a cost function based on the root mean square error (RMSE) criterion. The idea is to find the optimal set of parameter values that gives the minimum RMSE and produces the lowest model error. However, nonlinear model responses may produce multiple local minima in the cost function (cf. Posselt and Bishop, 2012), and thus sophisticated optimization algorithms are required to find the global minimum corresponding to the optimal parameters. Such optimization algorithms are usually too expensive computationally to be employed in sophisticated models. An alternative consists of supposing that the parameters are stochastic and of estimating them using filtering techniques (e.g. Annan and Hargreaves, 2004; Posselt and Bishop, 2012; Ruiz *et al.*, 2013; Schirber *et al.*, 2013). The basic idea to estimate the parameters is based on an augmented state composed of both the state of the system and the physical parameters in a nonlinear Gaussian state-space model. This online estimation is a tough problem in practice. Even a simple linear state equation with multiplicative parameters behaves nonlinearly for parameter estimation (Yang and Delsole, 2009).

Another approach consists of estimating the physical parameters independently of the state of the system. The particular advantage of using an offline estimation technique is that the control space is reduced from 10^7 to just a few dimensions. This drastic reduction in size permits us to conduct several model/parametrization evaluations as is often needed in parameter estimation. One disadvantage of offline techniques is that they cannot take into account the feedback of the changes that the parametrization produces onto the parametrization itself. Nevertheless, for a subgrid-scale orography (SSO) scheme, this issue should not be too critical, since most of the flow changes produced are advected downstream (for instance in the form of potential vorticity banners; Figure 13(c) in Lott, 1995).

Accordingly, the feedback can be neglected if the mountains considered are not close to the lee of other mountains.

In order to conduct an offline estimation of physical parameters, the parametrization should be compared to observations, for instance the Pyrenees Experiment (PYREX) campaign, in which surface drag and momentum fluxes were measured over a transect of the Pyrenees mountains (cf. Bougeault *et al.*, 1990). In this case, the mountain massif can be considered to be entirely located within a model gridbox area of a climate model, so drag and momentum flux can be directly compared to the same quantities predicted by the scheme over the same area. Therefore, we can validate SSO schemes using a single vertical column. This approach is often used prior to the implementation of the schemes in GCMs (e.g. the offline tests of the scheme in single vertical columns using the PYREX data in Lott and Miller, 1997).

Pulido and Thuburn (2005, 2008) showed that a four-dimensional variational data assimilation technique can be used to estimate the missing momentum forcing due to the unresolved/subgrid-scale gravity waves in the stratosphere. This missing momentum forcing was used to estimate optimal parameters of a non-orographic gravity wave parametrization in Pulido *et al.* (2012). Using twin experiments, they showed that the variational data assimilation technique does not converge towards the optimal parameters because of the nonlinear response of the parametrization to parameter perturbations. They employed a time-demanding genetic algorithm to overcome these difficulties. In the present work, we propose a similar offline parameter estimation procedure, but using a ensemble-based data assimilation technique to estimate the optimal parameters of a SSO scheme.

The technique presented here uses the Ensemble Kalman Filter (EnKF) and Ensemble Kalman Smoother (EnKS) which are reviewed in detail in Evensen (2009). In this work, we do not use an augmented state to estimate parameters of GCMs, as is usually done for online estimation (Annan and Hargreaves, 2007; Ruiz *et al.*, 2013); instead the state variables for the EnKF are only the physical parameters in this offline parameter estimation. As we do not have any knowledge of their temporal evolution, the state model is supposed to follow a random walk. In this way, we assume a non-negligible model error. An innovative part of our technique is that we also estimate the statistical parameters of the EnKF: (i) the covariance matrices of the Gaussian errors that control the weight of the state and the observation equations and (ii) the background state of the filter, typically an *a priori* knowledge of the physical parameters. Generally, these statistical parameters of the EnKF are prescribed values chosen by the user. In practice, this manual tuning does not ensure the filter convergence to the state of the system. To overcome this problem, the standard implementations of the EnKF use an inflation factor for the forecast and/or observational-error covariance matrices to avoid filter divergence. However, the main problem of this approach is the choice of the covariance inflation (additive or multiplicative) and the amplitude of the inflation. Several studies propose to estimate the inflation factors using the first moment estimation of the squared innovation (e.g. Wang and Bishop, 2003; Li *et al.*, 2009; Liang *et al.*, 2011), Bayesian approaches (e.g. Anderson, 2007; Miyoshi, 2011), or the second-order least-squares statistic of the squared innovation (Wu *et al.*, 2013). The technique presented in this article does not need to use any inflation factor since the statistical parameters are non-deterministic values. Here, as the estimation is offline in a low-dimensional system, we estimate directly the entire error covariance matrices and the background state of the EnKF using a maximum likelihood approach. In particular, we use the iterative and efficient Expectation–Maximization (EM) algorithm introduced by Dempster *et al.* (1977). To our knowledge, the implemented technique in this work based on the combination of an ensemble Kalman filter with the EM algorithm has not been proposed previously in data assimilation.

The novel estimation technique is applied to the SSO scheme described in Lott and Miller (1997) and revised in Lott (1999). This SSO scheme computes the wind tendencies due to the subgrid-scale orography and is implemented in three GCMs: that of the Laboratoire de Météorologie Dynamique (LMDz), the ECHAM model which is the atmospheric component of the Earth System Model of the Max Planck Institute (MPI-ESM), and that of the European Centre for Medium-Range Weather Forecasts (ECMWF). It is known that weather forecast and climate models are sensitive to the physical parameters of SSO schemes (e.g. Palmer *et al.*, 1986; Lott *et al.*, 2005; Sigmond *et al.*, 2008). Currently, this issue is still important since climate models now extend to the middle atmosphere where mountain gravity waves significantly affect the Brewer–Dobson circulation (McLandress and Shepherd, 2009). This circulation seems to intensify with climate change (Li *et al.*, 2008). These results call for a re-evaluation of the SSO schemes in the middle-atmosphere-resolving models and in particular of the set of parameters used in the schemes. An optimization of the SSO schemes can help to evaluate better the potential effects of the orographic gravity wave drag on the westerlies in midlatitudes.

This article is organized as follows. First, we describe the SSO scheme and the datasets it uses in section 2. Then, in section 3, we present the statistical model used to estimate the physical parameters of the SSO scheme. The details of the estimation technique based on the EnKF, EnKS and the EM algorithm are explained in section 4. The estimation technique is applied to a column version (not a 3D version) of the subgrid-scale orography scheme. We then use two synthetic cases (i.e. without using real observations): an identical-twin experiment and a situation in which the horizontal resolution of the orography dataset is changed. We show the results in section 5. Conclusions are drawn and future work is outlined in section 6. In general, the unified notations of data assimilation given in Ide *et al.* (1997) are used here.

2. Data and model

2.1. General circulation model data

To conduct our offline estimation we used daily data from a simulation done with the LMDz GCM (Hourdin *et al.*, 2006) using a horizontal resolution of $3.75^\circ \times 2.5^\circ$ and 50 vertical levels with a model top at 5 hPa. We have extracted from this model the SSO scheme we want to optimize. To conduct the optimization, we limit ourselves to a one-month period, July 2000. The exact year itself is of little importance, since the run considered has a spin-up of several years, and was not constrained by forcings other than the sea-surface temperature and the land–sea ice cover. The particular month chosen is in midwinter in the Southern Hemisphere, when high wind speed conditions prevail over the southern Andes.

The SSO scheme we use represents mountain gravity wave drag and blocked flow drag following Lott and Miller (1997). It also introduces lateral lift to take into account the fact that narrow valleys are partially sheltered from the large-scale winds in the free troposphere (cf. Lott, 1999). The scheme was extended to the stratosphere in Lott *et al.* (2005) and this is the version we use. For completeness, the salient features of scheme are described here.

Before launching a simulation, subgrid-scale orography parameters are calculated in each model gridbox: the mountain minimum, mean, and maximum elevations, and the mountain departure from the mean is then characterized by its anisotropy, its orientation angle, its slope and its standard deviation. As we will see, when we change orography datasets these parameters change significantly and the most dramatic changes concern the evaluation of the slope. We will adress these issues in section 5.2 and evaluate the changes to be done to the SSO scheme used in LMDz, when we make a transition from the 10 min of resolution US Navy orography dataset used in most current applications, to

Table 1. Physical parameters (defined in the text) of the SSO scheme, their assigned true values and their corresponding physical range.

Physical parameter	True value	Range
G	1	$(0, 1.5)$
C_d	1	$(0, 1.5)$
C_l	1	$(0, 1.5)$
H_{NC}	1	$(0, 1.5)$
β	0.5	$(0, 1)$
Ri_c	0.25	$(0, 2)$

a more refined 2 min of resolution dataset. At each time step, the SSO scheme uses the background flow conditions predicted by the model at a given gridpoint (i.e. the horizontal components of the winds, the temperature and the surface pressure), and predicts the effect of the SSO on the large-scale flow at all model levels.

The SSO scheme uses a set of six non-dimensional parameters of order O(1) which characterize the mesoscale and synoptic-scale effects of the mountain on the large-scale flow. The first three parameters directly scale the forces associated with the different processes parameterized: the gravity wave drag G, the low-level blocked flow drag C_d, and the low-level lift C_l that enhances large-scale vortex compression to represent valley sheltering. The other three parameters used are the low-level flow blocking depth H_{NC}, the fraction β of the gravity wave drag that propagates toward the free troposphere and aloft, and the critical Richardson number Ri_c that is used to predict when the mountain waves break. In Table 1, we give the values of each parameter used in operation, and also the range of values we will consider as plausible when we retune the SSO scheme.

The profiles of wind tendencies given by the SSO scheme are very sensitive to the value of the six non-dimensional parameters as in other schemes. The parameters used in the past were motivated by decades of research on mountain flow dynamics, and by a few experimental campaigns conducted over specific areas like the Pyrenees in France (e.g. Lott, 1995, gives a motivation for the lift based on the PYREX campaign). Although satellite data combined with high-resolution simulations could also be used in the future (Hertzog *et al.*, 2012), it remains the case that local tunings will probably still be needed, at least near places where the drag forces can potentially be very important. For this reason, and also because the methodology we propose is well adapted to handle one-column models (Posselt and Bishop, 2012), this is the strategy we have followed in the present work, where we imagine that an observational campaign takes place near the Perito Moreno Glacier in the Andes (46°S, 71°W) (dot in Figure 1). There, the mountains are characterized by an important anisotropic shape and strong variation in altitude (the standard deviation is 295 m for a mean altitude of 531 m and a peak of 1513 m). These topographical conditions represented in Figure 1(a) are ideal to study mountain-induced forces especially in high surface wind speed conditions such as in Figure 1(b). Indeed, this geographical location gave one of the largest subgrid-scale mountain drag amplitudes on the Earth in a preliminary spatial analysis for July 2000 in which we computed globally the subgrid-scale mountain drag with the scheme.

2.2. Preliminary tests

To give a preliminary view of the scheme outputs, Figure 2 shows the tendencies predicted by the SSO scheme,

$$\mathbf{y}(t_k) = \mathcal{F}\{\boldsymbol{\theta}, \mathbf{Z}(t_k)\}, \tag{1}$$

where $\boldsymbol{\theta} = (H_{NC}, C_d, Ri_c, G, C_l, \beta)$, and $\mathbf{Z}(t_k)$ is a generic notation for the vertical profiles of the horizontal winds and temperature. In Eq. (1), the vector $\mathbf{y}(t_k)$ has $m = 100$ values each of the 31 days $k \in \{1, \ldots, K = 31\}$ of July 2000. Each day, the first 50 values correspond to the zonal tendencies at the 50 model levels, and the last 50 values to the meridional tendencies.

From a preliminary temporal analysis during July 2000 at the chosen location, we distinguish two characteristic regimes of wind profiles in terms of the resulting induced SSO tendencies. Examples of these two regimes are shown in Figure 3. On 5 July 2000 (dotted line), the wind profile shows low and constant wind speeds with altitude. On 25 July (dashed line) the profile shows higher surface wind speeds and an increase of the wind with height, due to the presence of the subpolar jet in the region. The zonal and meridional components of the SSO tendency for the two cases are shown in Figure 2. The free physical parameters of the scheme are set to $H_{NC}^t = 1$, $C_d^t = 1$, $Ri_c^t = 0.25$, $G^t = 1$, $C_l^t = 1$ and $\beta^t = 0.5$. These set of 'true' parameter values $\boldsymbol{\theta}^t$ were proposed by Lott (1999) and generate the 'true' tendency denoted $\mathbf{y}^t(t_k)$. From Figure 2, on 25 July 2000, we find large tendencies whereas on 5 July 2000 the effect is much weaker due to the low wind speed conditions. The 95th percentile envelope around the mean value for the month of July 2000 is also shown. It indicates that the predicted SSO tendencies tend to be small in the mid-levels and larger at levels corresponding to the peak of the mountain (800 hPa) and to the tropospheric jet (250 hPa).

To evaluate how the outputs of the scheme vary with the different parameters, we evaluate the cost function

$$J(t_k) = \left\{\mathbf{y}^t(t_k) - \mathbf{y}(t_k)\right\}^\top \left\{\mathbf{y}^t(t_k) - \mathbf{y}(t_k)\right\}, \tag{2}$$

where the transpose notation $^\top$ is used, so that the square differences are summed over all altitudes and over the two components (zonal and meridional). Firstly, the cost function given in Eq. (2) is computed by changing one physical parameter and by fixing the other parameters to the true values.

Figure 4(a) shows the sensitivity of J as a function of H_{NC} and C_l for the state found on 5 July 2000 which is a situation with low surface winds as shown in Figure 3. The cost function associated with H_{NC} shows a non-quadratic behaviour, representing a nonlinear sensitivity in the derivative of J. The parameter C_l shows a quadratic cost function so that its sensitivity is linear. The other four parameters also show a linear sensitivity, as found for the C_l parameter, so that their cost functions are not shown. Figure 4(b) shows the sensitivity of J in high surface wind speed conditions, on 25 July 2000. The sensitivity of J to H_{NC} is enhanced by a factor of 10^3 in strong wind speed conditions compared to the sensitivity for the weak wind speed case. Finally, a relatively weaker enhancement of the J sensitivity is found for high surface wind speed conditions to certain physical parameters, e.g. C_l (squares) and C_d (not shown here) compared to the enhancement of H_{NC} sensitivity between low and high surface wind conditions. In Figure 4(b), a saturation of the cost function is found close to the global minimum for $H_{NC} > 1.1$ (circles) in high wind speed conditions. This behaviour can be explained as follows. As the surface wind increases and H_{NC} increases, the blocked flow depth decreases and eventually reaches 0 (Eqs. (4) and (9) in Lott and Miller, 1997). At this point, the parameter H_{NC} becomes saturated since an increase of its value cannot change the blocked flow depth to negative values. Therefore, H_{NC} values larger than this critical value cannot affect the SSO predictions.

In a second sensitivity experiment, the cost function given in Eq. (2) is computed changing two physical parameters simultaneously. Figure 5 shows the cost function as a function of the parameters H_{NC} and G. H_{NC} is correlated with G. The ten smallest values of the cost function are indicated by black dots in Figure 5. They underline the fact that the global minimum region of the cost function (intersection of the two dashed black lines) is not well defined. On the contrary, in Figure 5(a), a large region of very low sensitivity close to the global minimum is highlighted, especially in low wind speed conditions where the sensitivity of J is reduced. In this region of the cost function, there is a negative correlation between H_{NC} and G.

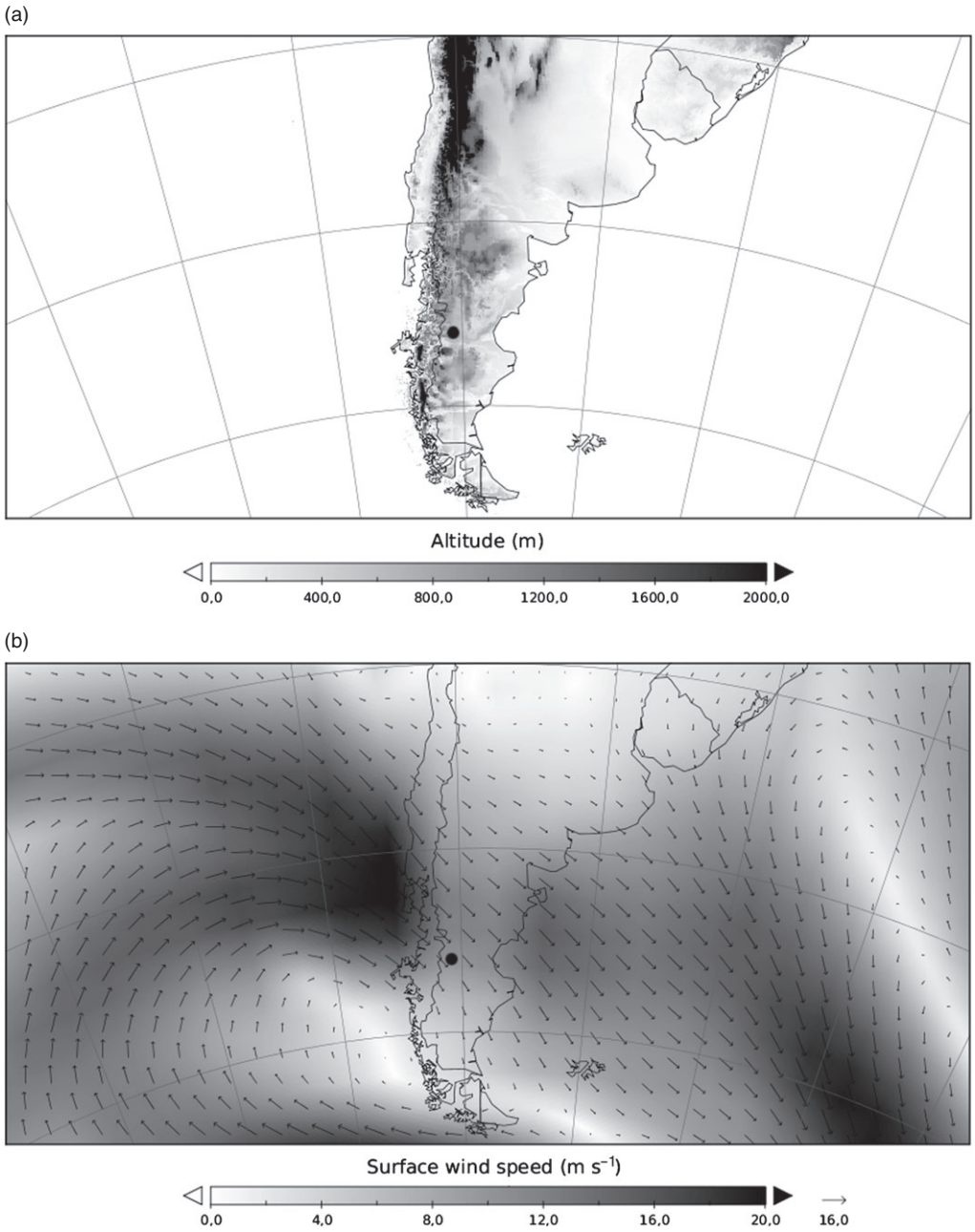

Figure 1. (a) Topography of the south Andes and (b) surface winds on 25 July 2000. The location of the chosen mountain peaks is close to the Perito Moreno Glacier (46°S, 71°W) and represented by a dot.

3. Nonlinear Gaussian state-space model

To estimate the $n = 6$ physical parameters in $\boldsymbol{\theta}$ via our filtering technique, we first need to make them stochastic. We denote them as $\mathbf{x}$ and we say it is the 'state of the system'. The state evolution is given by a Gaussian random walk,

$$\mathbf{x}(t_k) = \mathbf{x}(t_{k-1}) + \boldsymbol{\eta}(t_k), \tag{3}$$

where the n-dimensional stochastic random vector $\{\boldsymbol{\eta}(t_k)\}_{k\in\{1,\ldots,K\}}$ represents an additive perturbation at each time t_k. We assume that the perturbations are Gaussianly distributed with zero mean and a constant in time $n \times n$ covariance matrix $\mathbf{Q}$. Equation (3) is taken as the state equation in our state-space model.

If we use directly the physical parameters $\boldsymbol{\theta}$ as the state of the system, they can easily become negative or reach very large values, whereas the parameters in the SSO scheme are assumed to be always positive and of the order of unity. For this reason, we map the physical parameters $\boldsymbol{\theta}$ on $\mathbf{x}$ by using the Gauss error function $\boldsymbol{\theta} = \mathcal{G}(\mathbf{x})$, as sometimes used in data assimilation (Hu *et al.*, 2010).

At the initial time of Eq. (3), we introduce an *a priori* knowledge of the physical parameters. We assume that this background information follows a Gaussian distribution given by the n-dimensional vector mean $\mathbf{x}^{\mathrm{b}}$ and the $n \times n$ covariance matrix $\mathbf{B}$.

At time t_k, the zonal and meridional SSO tendencies given in Eq. (1) are assumed to be observed. They are stored in the m-dimensional stochastic random vector $\{\mathbf{y}(t_k)\}_{k\in\{1,\ldots,K\}}$. The state vector at time t_k is related to the observation by means of the observation equation defined by

$$\mathbf{y}(t_k) = \mathcal{H}_k\left\{\mathbf{x}(t_k)\right\} + \boldsymbol{\epsilon}(t_k), \tag{4}$$

where the observation operator $\mathcal{H}_k$ is the nonlinear function defined by

$$\mathcal{H}_k\left\{\mathbf{x}(t_k)\right\} = \mathcal{F}\left[\mathcal{G}\left\{\mathbf{x}(t_k)\right\}, \mathbf{Z}(t_k)\right], \tag{5}$$

where $\mathcal{F}$ is the SSO scheme Eq. (1) and $\mathcal{G}$ is the Gauss error function. In Eq. (4), we suppose that the m-dimensional stochastic random vector $\{\boldsymbol{\epsilon}(t_k)\}_{k\in\{1,\ldots,K\}}$ is an additive zero-mean Gaussian error. The $m \times m$ covariance matrix of $\boldsymbol{\epsilon}(t_k)$ is denoted by $\mathbf{R}(t_k)$.

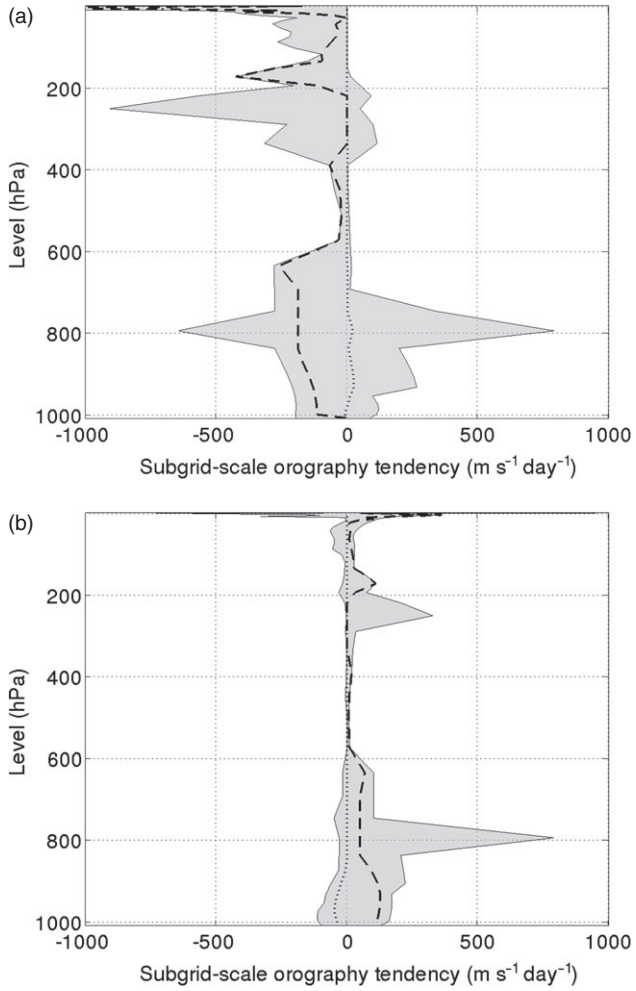

Figure 2. Vertical profiles of (a) zonal and (b) meridional tendencies generated by the SSO scheme at location 46°S, 71°W within the Andes. The grey shading denotes the 95th percentile envelope over the month of July 2000. The dotted and dashed lines correspond to the mountain drag profiles on 5 and 25 July 2000 respectively.

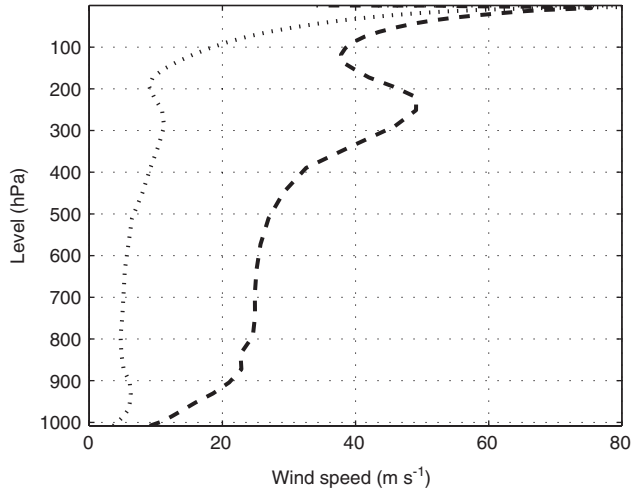

Figure 3. Wind speed profiles from the LMDz GCM model on 5 (dotted line) and 25 (dashed line) July 2000 at location 46°S, 71°W within the Andes.

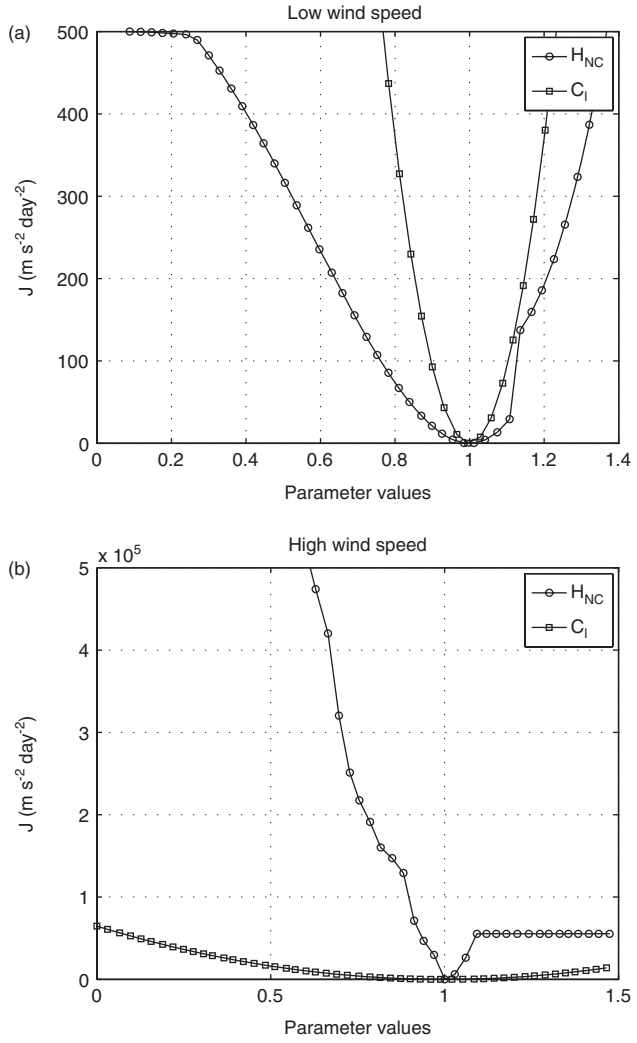

Figure 4. Cost function (Eq. 2)) as a function of the physical parameters H_{NC} (circles) and C_{l} (squares). The true values of the physical parameters are $H_{\mathrm{NC}}^{\mathrm{t}} = 1$ and $C_{\mathrm{l}}^{\mathrm{t}} = 1$. The results are given for the location 46°S, 71°W on (a) 5 July 2000 representing low surface wind speed conditions and (b) 25 July 2000 representing high surface wind speed conditions.

As the sensitivity of J varies with the atmospheric conditions $\mathbf{Z}(t_k)$, particularly with the surface wind speed, $\mathbf{R}(t_k)$ is assumed in principle to vary with time.

The statistical parameters correspond to the vector and matrices that define the system (3) and (4). They are denoted by $\boldsymbol{\psi}$. We use the term 'statistical parameters' of the state-space statistical model to distinguish from the six 'physical parameters' $\boldsymbol{\theta}$ of the SSO scheme. The statistical parameters are the *a priori* probability density functions (PDFs) of the physical parameters, given by $\mathbf{x}^{\mathrm{b}}$ and $\mathbf{B}$, and the covariance error matrices $\mathbf{Q}$ and $\mathbf{R}(t_k)\ \forall k \in \{1, ..., K\}$. We write $\boldsymbol{\psi} = \left(\mathbf{x}^{\mathrm{b}}, \mathbf{B}, \mathbf{Q}, \mathbf{R}\right)$. The statistical parameters define the uncertainty of the state-space statistical model and play a central role on the quality and rate of convergence in the estimation of the physical parameters with the filtering and smoothing techniques described below.

The estimation of the statistical parameters $\boldsymbol{\psi}$ is conducted maximizing the total likelihood function $\mathcal{L}$. This function is based on the PDF of the initial state $p\{\mathbf{x}(t_1)\}$, the conditional state evolution $p\{\mathbf{x}(t_k)|\mathbf{x}(t_{k-1})\}$ and the observations conditionally to the state $p\{\mathbf{y}(t_k)|\mathbf{x}(t_k)\}$. The three PDFs are assumed to be normally distributed with the respective mean and covariances: $\mathbf{x}(t_1) - \mathbf{x}^{\mathrm{b}}$ and $\mathbf{B}$, $\mathbf{x}(t_k) - \mathbf{x}(t_{k-1})$ and $\mathbf{Q}$, $\mathbf{y}(t_k) - \mathcal{H}_k\{\mathbf{x}(t_k)\}$ and $\mathbf{R}(t_k)$. Finally, using the Markov property of the state-space model, the total likelihood function is the product of the PDF for all times K. It is given by

$$\mathcal{L}(\mathbf{x}, \boldsymbol{\psi}) = p\{\mathbf{x}(t_1)\} \prod_{k=2}^{K} p\{\mathbf{x}(t_k)|\mathbf{x}(t_{k-1})\} \times \prod_{k=1}^{K} p\{\mathbf{y}(t_k)|\mathbf{x}(t_k)\}. \tag{6}$$

In practice, this total likelihood function is approximated by its expectation conditionally to all the observations

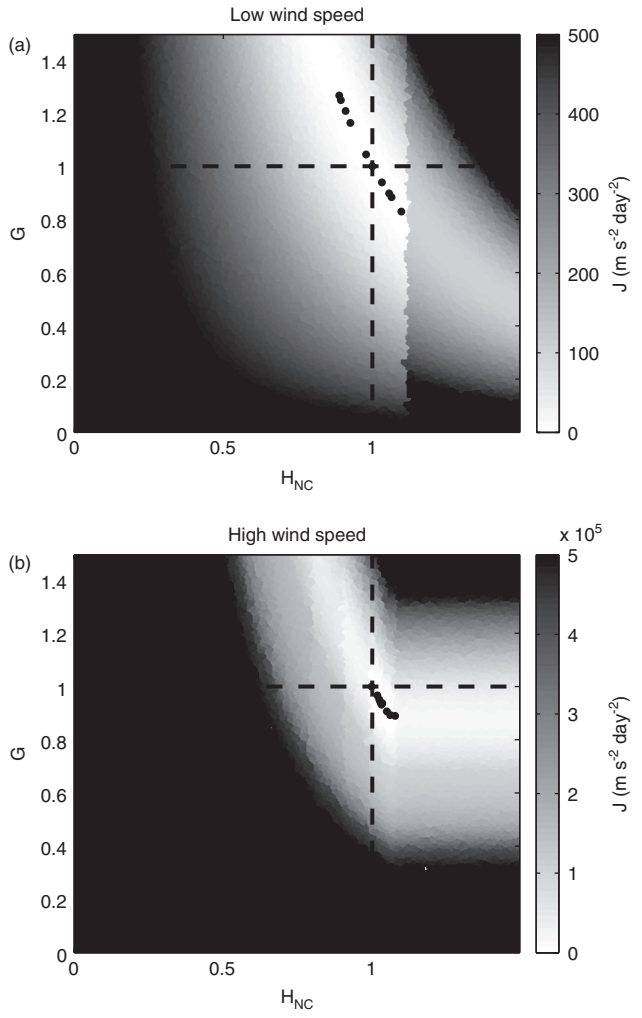

Figure 5. Cost function (Eq. (2)) as a function of the physical parameters H_{NC} (x-axis) and G (y-axis). The intersection of dashed black lines corresponds to the true physical parameters $H_{\mathrm{NC}}^{\mathrm{t}} = 1$ and $G^{\mathrm{t}} = 1$. The black dots denote the physical parameters which generate the ten lowest values of the cost function. The results are given for the location 46°S, 71°W on (a) 5 July 2000 representing low surface wind speed conditions and (b) 25 July 2000 representing high surface wind speed conditions.

$\mathbf{y}_{1:K} = \mathbf{y}(t_1), \ldots, \mathbf{y}(t_K)$. This requires the computation of the state smoothed probabilities to be described below.

4. Estimation technique

The algorithm to estimate the physical and statistical parameters is described concisely here. A diagram with the main steps of the algorithm is shown in Figure 6. The algorithm starts with a proposed set of statistical parameters $\widehat{\boldsymbol{\psi}}^{(1)}$ which do not need to be known precisely. Then, the statistical parameters are estimated by maximizing the total likelihood function using the EM algorithm. A loop is initiated which is composed by an expectation and a maximization step. The expectation step computes the expectations given in Appendix A via the EnKS. The maximization step consists basically in computing the optimal $\widehat{\boldsymbol{\psi}}^{(j)}$ from the known analytical expressions given in Appendix B. At each iteration j of the EM algorithm, we compute the innovation likelihood l given in Appendix C. It is commonly used to evaluate the quality of the state estimates and to compare state-space models with different statistical parameters (Cappé *et al.*, 2005, p. 140 give more details). If the innovation likelihood does not change significantly, the last estimated $\widehat{\boldsymbol{\psi}}^{(j)}$ is returned. These optimal statistical parameters given by the EM algorithm are finally used to initiate a last EnKF run which estimates the physical parameters.

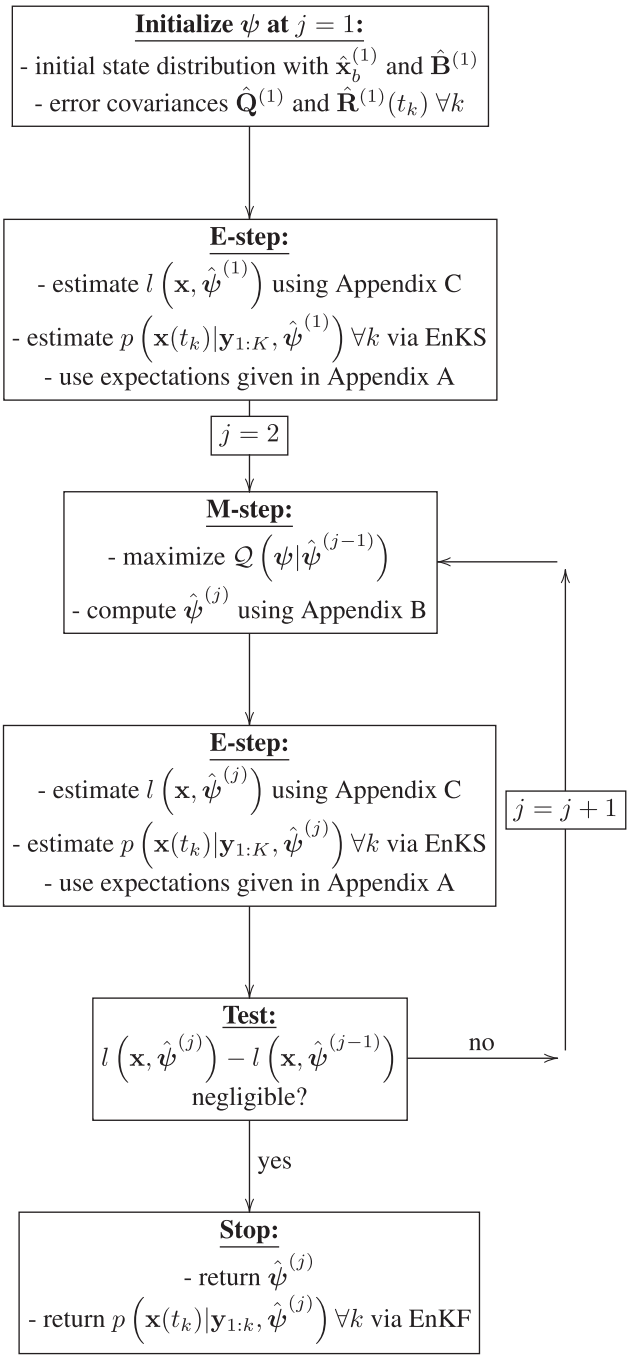

Figure 6. Diagram of the method based on the maximum likelihood estimates of the state-space model Eqs (3) and (4).

4.1. *Expectation-maximization algorithm*

The maximum likelihood estimates of the statistical parameters $\boldsymbol{\psi}$ are conducted using the EM algorithm proposed by Dempster *et al.* (1977). This is a classical method used in the case of incomplete or missing data. This iterative algorithm is based on two steps: the expectation of the total log-likelihood function (E step) and its maximization with respect to $\boldsymbol{\psi}$ (M step). The EM algorithm begins with an initial set of statistical parameters $\widehat{\boldsymbol{\psi}}^{(1)}$. Then, repeating the E and M steps, the sequence of estimates $\widehat{\boldsymbol{\psi}}^{(j)}$ yields increasing values of the expected log-likelihood and converges to the maximum likelihood estimates.

At iteration j, the E step consists of computing the expected total log-likelihood function conditionally to the total observations and the previously estimated statistical parameters. It is given by

$$\mathcal{Q}(\boldsymbol{\psi}|\widehat{\boldsymbol{\psi}}^{(j-1)}) = E\left[\log\{\mathcal{L}(\mathbf{x}, \boldsymbol{\psi})\} \mid \mathbf{y}_{1:K}, \widehat{\boldsymbol{\psi}}^{(j-1)}\right]. \quad (7)$$

In the case of nonlinear state-space statistical models, the exact smoothed probabilities are not computable. Thus, we use the Monte Carlo approximations given by the EnKS. The conditional expectations given in Appendix A are then computed.

The M step consists of maximizing $\mathcal{Q}(\boldsymbol{\psi}|\widehat{\boldsymbol{\psi}}^{(j-1)})$ with respect to $\boldsymbol{\psi}$. We obtain a direct analytic form of the maximum likelihood estimates. The expressions are given in Appendix B. The derivations are not presented here (cf. Tandeo *et al.*, 2011, for more details).

4.2. Ensemble Kalman filter

The EnKF algorithm used here is an adaptation of the one proposed by Burgers *et al.* (1998).

In the initial step of the EnKF algorithm, at time t_1, an ensemble of $\mathbf{x}$s composed of N members is randomly generated. The members of the ensemble follow a Gaussian distribution given by the vector mean $\mathbf{x}^{\mathrm{b}}$ and the covariance matrix $\mathbf{B}$. The N initial members are stored in the vectors $\mathbf{x}_i^{\mathrm{f}}(t_1)\ \forall i \in \{1, ..., N\}$.

In the update step, at each time t_k, we randomly generate N samples of $\boldsymbol{\eta}_i$ and $\boldsymbol{\epsilon}_i\ \forall i \in \{1, ..., N\}$ with respective covariances $\mathbf{Q}$ and $\mathbf{R}(t_k)$. Then, following Eq. (3), the i-member of the updated state is given by

$$\mathbf{x}_i^{\mathrm{f}}(t_k) = \mathbf{x}_i^{\mathrm{a}}(t_{k-1}) + \boldsymbol{\eta}_i(t_k), \tag{8}$$

and the mapping from the forecast state space to the observational space of the i-member is computed as

$$\mathbf{y}_i^{\mathrm{f}}(t_k) = \mathcal{H}_k\left\{\mathbf{x}_i^{\mathrm{f}}(t_k)\right\}. \tag{9}$$

The N members of the ensemble are used to estimate the sample means of the propagated state in the state space and in the observational space denoted by $\mathbf{x}^{\mathrm{f}}(t_k)$ and $\mathbf{y}^{\mathrm{f}}(t_k)$ respectively.

In the analysis step, we follow the Pham (2001) methodology which avoids the linearization of the observational operator. The Kalman gain is computed with

$$\mathbf{K}(t_k) = \mathbf{P}_{xy}^{\mathrm{f}}(t_k)\left\{\mathbf{P}_{yy}^{\mathrm{f}}(t_k) + \mathbf{R}(t_k)\right\}^{-1}, \tag{10}$$

where $\mathbf{P}_{xy}^{\mathrm{f}}(t_k)$ is the sample cross-covariance matrix and $\mathbf{P}_{yy}^{\mathrm{f}}(t_k)$ is the sample covariance matrix, which are determined by

$$\mathbf{P}_{xy}^{\mathrm{f}}(t_k) = \frac{1}{N-1}\sum_{i=1}^{N}\left\{\mathbf{x}_i^{\mathrm{f}}(t_k) - \mathbf{x}^{\mathrm{f}}(t_k)\right\}\left\{\mathbf{y}_i^{\mathrm{f}}(t_k) - \mathbf{y}^{\mathrm{f}}(t_k)\right\}^{\top} \tag{11}$$

and

$$\mathbf{P}_{yy}^{\mathrm{f}}(t_k) = \frac{1}{N-1}\sum_{i=1}^{N}\left\{\mathbf{y}_i^{\mathrm{f}}(t_k) - \mathbf{y}^{\mathrm{f}}(t_k)\right\}\left\{\mathbf{y}_i^{\mathrm{f}}(t_k) - \mathbf{y}^{\mathrm{f}}(t_k)\right\}^{\top}. \tag{12}$$

In our case, the number of observations is larger than the dimension of the state space ($m > n$), so that the matrix $\mathbf{P}_{yy}^{\mathrm{f}}(t_k) + \mathbf{R}(t_k)$ is ill-conditioned making the matrix inversion difficult. Therefore, as described in Evensen (2009), Chapter 14, we compute the pseudo-inverse of $\mathbf{P}_{yy}^{\mathrm{f}}(t_k) + \mathbf{R}(t_k)$, taking into account 99% of the information given by the eigenvalues. Having $\mathbf{K}(t_k)$ from Eq. (10), the N members of the ensemble are then updated by

$$\mathbf{x}_i^{\mathrm{a}}(t_k) = \mathbf{x}_i^{\mathrm{f}}(t_k) + \mathbf{K}(t_k)\mathbf{d}_i(t_k), \tag{13}$$

where the m-dimensional $\mathbf{d}_i(t_k)\ \ \forall i \in \{1, ..., N\}$ are the N innovation vectors in which we use perturbed observations such as $\mathbf{d}_i(t_k) = \mathbf{y}(t_k) + \boldsymbol{\epsilon}_i(t_k) - \mathbf{y}_i^{\mathrm{f}}(t_k)$. Note that the sample covariance of the N innovations is $\mathbf{P}_{yy}^{\mathrm{f}}(t_k) + \mathbf{R}(t_k)$. Finally, the updated analyzed state is represented by the sample mean $\mathbf{x}^{\mathrm{a}}(t_k)$ and the sample covariance $\mathbf{P}^{\mathrm{a}}(t_k)$.

4.3. Ensemble Kalman smoother

The backward recursions correspond to the EnKS algorithm proposed by Evensen and Van Leeuwen (2000). It uses the results of the EnKF computed above.

In the initial step of the EnKS algorithm, at time t_K, we use the members of the filtered state, $\forall i \in \{1, ..., N\}$, such as $\mathbf{x}_i^{\mathrm{s}}(t_K) = \mathbf{x}_i^{\mathrm{a}}(t_K)$ and $\mathbf{P}^{\mathrm{s}}(t_K) = \mathbf{P}^{\mathrm{a}}(t_K)$.

Then, we proceed backward from $k = K - 1$ to $k = 1$. At each time t_k, we compute

$$\mathbf{x}_i^{\mathrm{s}}(t_k) = \mathbf{x}_i^{\mathrm{a}}(t_k) + \mathbf{K}^{\mathrm{s}}(t_k)\left\{\mathbf{x}_i^{\mathrm{s}}(t_{k+1}) - \mathbf{x}_i^{\mathrm{f}}(t_{k+1})\right\}, \tag{14}$$

where $\mathbf{K}^{\mathrm{s}}(t_k)$ is the $n \times n$ Kalman smoother gain matrix given by $\mathbf{P}^{\mathrm{a}}(t_k)\left\{\mathbf{P}^{\mathrm{f}}(t_{k+1})\right\}^{-1}$. The Gaussian distribution of the updated state estimate is given by the sample mean and covariance respectively denoted by $\mathbf{x}^{\mathrm{s}}(t_k)$ and $\mathbf{P}^{\mathrm{s}}(t_k)$. The sample covariance of the state between two consecutive times is computed using

$$\mathbf{P}^{\mathrm{s}}(t_k, t_{k-1}) = \frac{1}{N-1}\sum_{i=1}^{N}\left\{\mathbf{x}_i^{\mathrm{s}}(t_k) - \mathbf{x}^{\mathrm{s}}(t_k)\right\}\left\{\mathbf{x}_i^{\mathrm{s}}(t_{k-1}) - \mathbf{x}^{\mathrm{s}}(t_{k-1})\right\}^{\top}. \tag{15}$$

5. Results

5.1. Identical-twin experiment

In order to evaluate the technique, twin experiments are used. In this case, the observations are obtained under the assumption of a perfect model; in other words, the SSO scheme is assumed to give the true tendencies when the physical parameters $\boldsymbol{\theta}^{\mathrm{t}} = (1, 1, 0.25, 1, 1, 0.5)$ are used as the true parameters. Then we suppose that the state, i.e. the physical parameters, is unknown and we try to estimate it via the state-space model (3) and (4) using the generated mountain drag observations. As schematized in Figure 6, we estimate the statistical parameters $\boldsymbol{\psi} = \left(\mathbf{x}^{\mathrm{b}}, \mathbf{B}, \mathbf{Q}, \mathbf{R}\right)$ of the state-space model via the EM algorithm in order to improve the estimation of the physical parameters, $\boldsymbol{\theta}$. At iteration $j = 1$, we deliberately initialize the state vector $\widehat{\mathbf{x}}_{\mathrm{b}}^{(1)}$ far from the true state values (corresponding to the true parameters). The corresponding covariance $\widehat{\mathbf{B}}^{(1)}$ is chosen as the unit matrix $\mathbf{I}_6$ to generate large initial spreads of the members. Throughout the filter evolution, the members are randomly perturbed by the constant covariance matrix $\widehat{\mathbf{Q}}^{(1)} = 0.1 \times \mathbf{I}_6$ in Eq. (8). The covariance of the measurement errors in Eq. (9) is set to $\widehat{\mathbf{R}}^{(1)}(t_k) = 1000 \times \mathbf{I}_{100}\ \ \forall k \in \{1, \ldots, K\}$, which is of the same order as the mean value of the cost function J given in Eq. (2). We use $N = 100$ members and 25 iterations of the EM algorithm.

The innovation log-likelihood function and the total RMSE of the physical parameters for the conducted twin experiments are shown in Figure 7(a) as a function of the EM iteration. The results indicate that the innovation log-likelihood is a good synthetic indicator of the filter quality which follows the inverse variations of the total RMSE. In Figure 7(b), we decompose the total RMSE for each physical parameter. We find a good convergence of all the physical parameters after $j = 10$ iterations except for C_{d}, H_{NC} and G, which need more EM iterations. The evolution of these two last physical parameters as a function of time for different iterations ($j = 1, 10, 25$) of the EM algorithm are shown in Figure 8. For both physical parameters, the EM algorithm is able to adapt the filter conditions and to give, along the iterations j, more and more accurate initial distributions of the physical parameters (given by the $\mathbf{x}^{\mathrm{b}}$ and $\mathbf{B}$ maximum likelihood estimates). However, at the last iteration $j = 25$, the temporal convergence (near $k = 20$) is higher than the other physical parameters (not shown here). Note that the results using deterministic values of $\boldsymbol{\psi}$, instead of estimating them via the maximum likelihood method, show the inability of the filter to converge to the solution $\boldsymbol{\theta}^{\mathrm{t}}$. This is shown with the

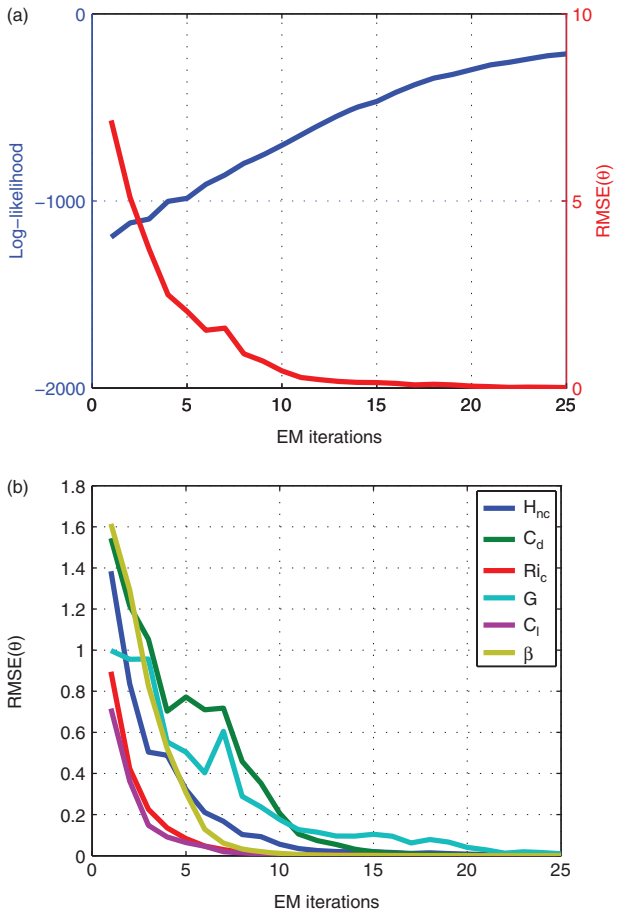

Figure 7. Evolution of (a) the innovation log-likelihood (left y-axis; blue line), the total physical parameter RMSE summed by time (right y-axis; red line) and (b) detailed RMSE of each physical parameter over $j = 25$ iterations of the EM algorithm.

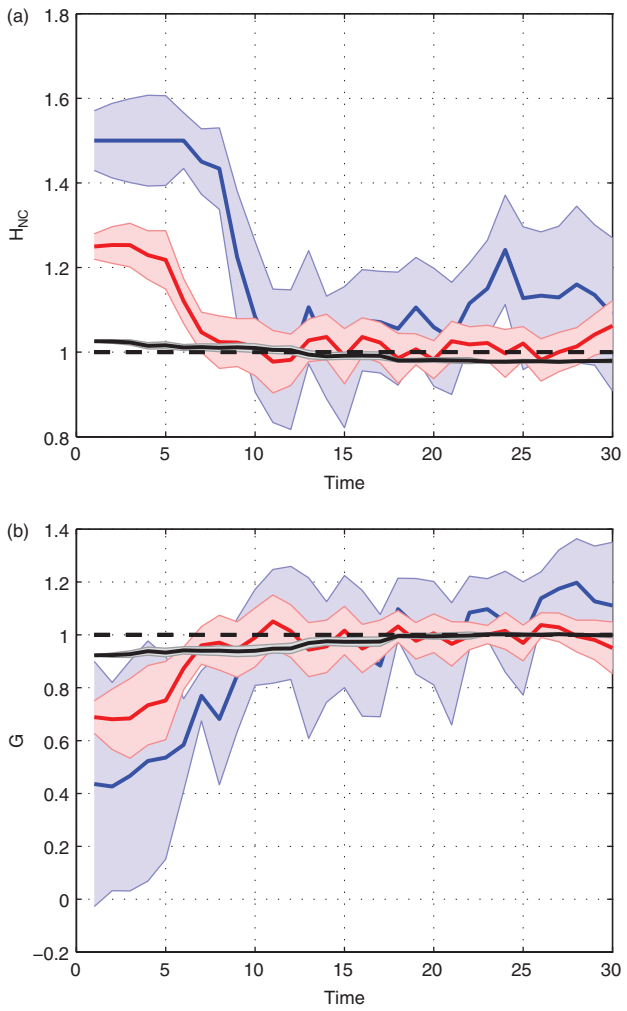

Figure 8. Evolution of the (a) H_{NC} and (b) G estimates with their 95% confidence intervals over time for different iterations of the EM algorithm: $j = 1$ (blue), $j = 10$ (red) and $j = 25$ (black). The straight lines correspond to the true physical parameter values. The physical parameters ($\boldsymbol{\theta}$, not $\mathbf{x}$) are shown.

blue curves of Figure 8 corresponding to the first iteration of the EM algorithm (i.e. this could be interpreted as a standard EnKF estimation). Even if we use more realistic but uniform values of covariance matrices $\widehat{\mathbf{B}}^{(1)}$, $\widehat{\mathbf{Q}}^{(1)}$ and $\widehat{\mathbf{R}}^{(1)}(t_k)$ $\forall k \in \{1, \ldots, K\}$, the standard EnKF is unable to converge to a stable and accurate solution.

Figure 9(a) shows the matrix $\mathbf{Q}$ after 25 EM iterations. A negative correlation between the H_{NC} and G physical parameters is clearly detected. This confirms the observation we made from Figure 5 in the weak sensitivity region of the cost function J. The elements of the $\mathbf{Q}$ estimated by the maximum likelihood method for the covariance between H_{NC} and G and the variances of H_{NC} and G are respectively -1.5×10^{-5}, 1.5×10^{-5} and 3×10^{-5}. These variances correspond to the optimal perturbations of the members in Eq. (3) at each time of the filter. Note that the amplitude of $\mathbf{Q}$ tends to decrease with the iterations of the EM algorithm since the model becomes perfect and the observations are produced with the optimal physical parameters. Concerning the estimated amplitude of the observation-error covariance $\mathbf{R}(t_k)$, i.e. the covariance of $\boldsymbol{\epsilon}(t_k)$ $\forall k \in \{1, \ldots, K\}$, it varies with the forcing terms, particularly the surface wind speed conditions. The results for the low and high wind speed conditions are shown in Figure 9(b)and (c). We distinguish different parts on these estimated matrices. The top left and the bottom right parts correspond respectively to the zonal and meridional error covariances of the observation equation given in Eq. (4). The top right and bottom left parts correspond to the cross-covariance between the zonal and meridional components. The x- and y-axes indicate the vertical level of the different components. For instance, the level 1000 hPa is given by the indices 1 and 51 whereas the level 5 hPa is given by the indices 50 and 100. The results indicate a checkerboard structure in the covariances inside groups of vertical levels and especially a larger variability of the observation error in the levels close to the surface in both cases. We also observe a cross-covariance between the zonal and meridional error terms at this altitude for the low wind speed conditions. The main difference between the two estimated matrices is the amplitude of the variability. In strong wind speed conditions, the variance is globally enhanced by a factor of 10^3. Therefore, the method proposed here is able to model a flow-dependent (typically the wind speed) and not necessarily diagonal error covariance matrix $\mathbf{R}$. Miyoshi *et al.* (2012) have also proposed to retrieve the shape of $\mathbf{R}$ in a data assimilation problem conducting twin experiments. More precisely, they extended the adaptive estimation method proposed by Li *et al.* (2009) to include off-diagonal terms of $\mathbf{R}$.

We make two comments on results that are not shown here. Firstly, the use of $N = 500$, 1000 members in the ensemble (not shown) gives similar results as the case with $N = 100$ presented here. Thus, an ensemble of 100 members is sufficient to capture the highly nonlinear behaviour of the SSO scheme and to estimate properly the statistical parameters of the state-space system. Secondly, the maximum likelihood statistical parameter $\boldsymbol{\psi}$ estimates are independent of the initial conditions of the EM algorithm. Different initial guess parameters $\widehat{\mathbf{x}}_{\mathrm{b}}^{(1)}$ and different covariances $\widehat{\mathbf{B}}^{(1)}$ give similar rates of convergence.

5.2. *Changes in orography resolution*

When the resolution of a GCM is increased, or when a new dataset is used to feed the physical parametrizations, the physical parameters of the GCM need to be adjusted. There is no systematic

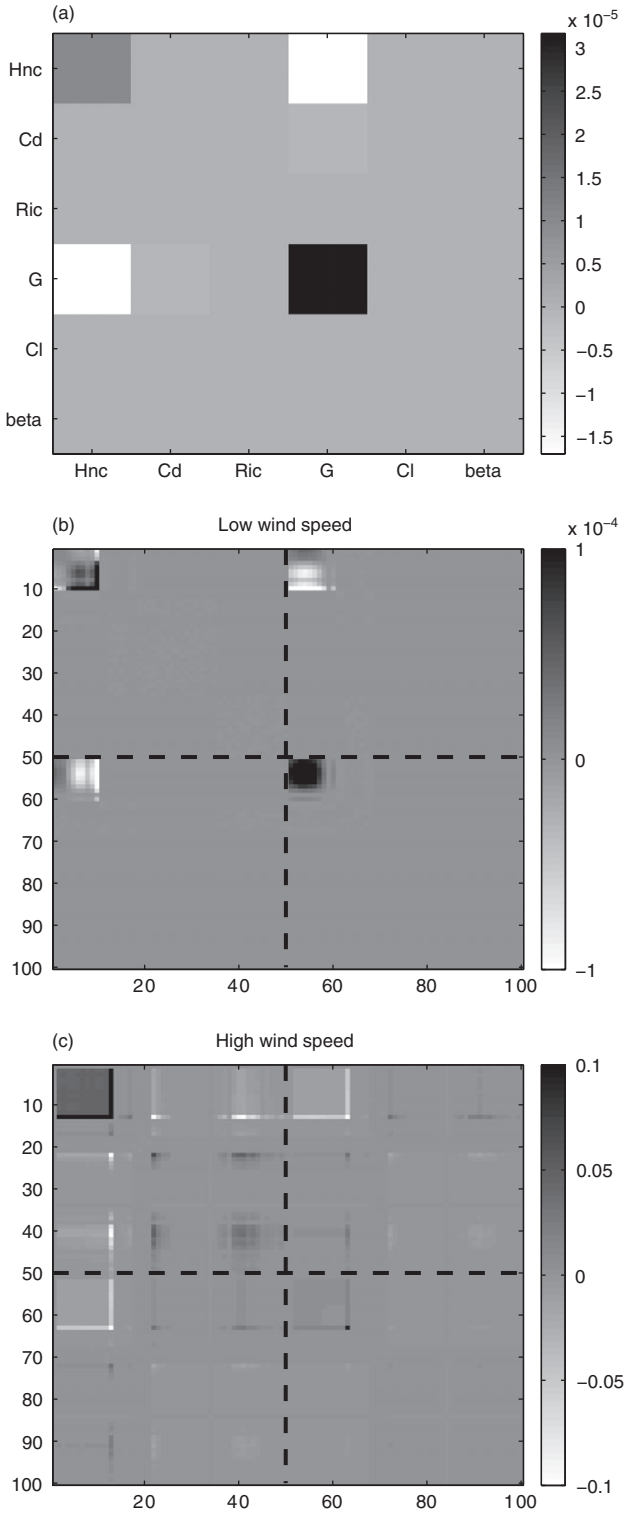

Figure 9. Maximum likelihood estimates after $j = 25$ iterations of the EM algorithm of (a) $\widehat{\mathbf{Q}}$, (b) $\widehat{\mathbf{R}}(t_5)$ and (c) $\widehat{\mathbf{R}}(t_{25})$. The straight lines in (b) and (c) denote the limit between the zonal and meridional mountain drag error covariance of the observation equation.

way to produce these adjustments in the schemes so far. The technique introduced in this work can be used to do this. In particular, the standard parameters that are currently used in the SSO scheme shown in Table 1 have been manually tuned using PYREX data by Lott and Miller (1997). This set of parameters are used operationally in the LMDz model. The tuning was conducted with a version of the SSO scheme that uses the low-resolution orography ($10' \times 10'$; Figure 10(a)). Suppose that the higher resolution ($2' \times 2'$; Figure 10(b) and NOAA, 2001) orography dataset is used to improve LMDz at a given horizontal resolution. The parameters of the scheme should be adjusted for this new orography dataset. We conducted an experiment to examine if the technique is able to determine a new set of optimal parameters for this high-resolution orography dataset. The conducted data assimilation experiments use the SSO tendencies predicted by the SSO scheme using the low-resolution orography as observations. The assimilation then uses the SSO tendencies predicted with a higher-resolution orography dataset. In this way, the operator $\mathcal{H}_k$ used in the assimilation has an error.

In this experiment in which the SSO scheme is not 'perfect' due to the resolution change, we found that the results depend on the initial guess conditions, in opposition to the identical-twin experiments. As convergence could not be reached easily, one hundred filter experiments with different random initial guess conditions $\widehat{\mathbf{x}}_{\mathrm{b}}^{(1)}$ of the EM algorithm were conducted. Among these 100 experiments, we find results with recurrent estimations reach the same log-likelihood as those shown in Figure 11(a). The parameter estimations after $j = 25$ EM iterations for five selected cases are shown in Figure 11(b).

From Figure 11(b), we notice that there is one parameter which does not need to be changed much when the resolution is changed $-C_{\mathrm{l}}$. This is not a surprise since C_{l} is an almost linear lift coefficient, which is related to a mountain lift force whose amplitude varies linearly with the difference between the mountain and valley height. We also find that C_{d} needs to be reduced by a factor ≈ 2. Considering Eq. (16) in Lott and Miller (1997), the scheme measures the number of mountains in a subgrid-scale area, and multiplies the low-level drag by this number of ridges. This yields a multiplicative factor in the mountain slope. When we move to a higher-resolution grid, the estimate of the slope necessarily increases, so C_{d} needs to decrease. The same conclusion could be drawn for the parameter G which controls the gravity wave drag, but here the technique gives two possible solutions. One where G is almost unchanged or has a weak increase, and one where it is decreased substantially, as expected. As the solution with unchanged G is the most surprising, it is important to notice that this is also related to a smaller β; they therefore correspond to more trapped waves which apply more low-level drag. As at the low level, it is C_{d} that essentially controls the drag. We have therefore increased the gravity wave drag by increasing G but placed that drag at low level where the effect is small compared to that of C_{d}. Another important result of the analysis is that the value of the critical Richardson number clearly converges to $Ri_{\mathrm{c}} = 1.5$. As this high-resolution orography case likely has larger-amplitude gravity waves, this larger Richardson number than the one used with the low-resolution orography dataset needs to be enhanced so that the waves propagate at high levels without breaking systematically at lower levels.

In general, the parameter estimations, except for Ri_{c}, present a very large spread, particularly for those parameters acting at low levels. For these, it should be remembered that the drag at low levels is always treated via implicit methods in part for stability, and in part because overestimated drags could yield wind reversals at low levels, which contradict the nature of drag forces. Clearly, the assimilation technique indicates that some physical considerations should be given to make these parameters more efficient in controlling the drag. Among the possibilities, the SSO scheme does not consider that, when there are several mountains in a gridbox area, some sheltering should be taken into account not to decelerate the same flow twice in succession. This is currently handled implicitly by the scheme, but the low-level drag should take into account this horizontal sheltering when we increase the orography resolution.

Figure 11(b) shows that the filter converges towards two possible optimal states, in which H_{NC} and G clearly present bimodal distributions. This result is associated with the high correlation that was found in the cost function between H_{NC} and G (Figure 5). The presence of model error in this imperfect model experiment appears to add complexity to the cost function with the presence of these two local minima. This is consistent with the results obtained by Schirber *et al.* (2013) in an online parameter

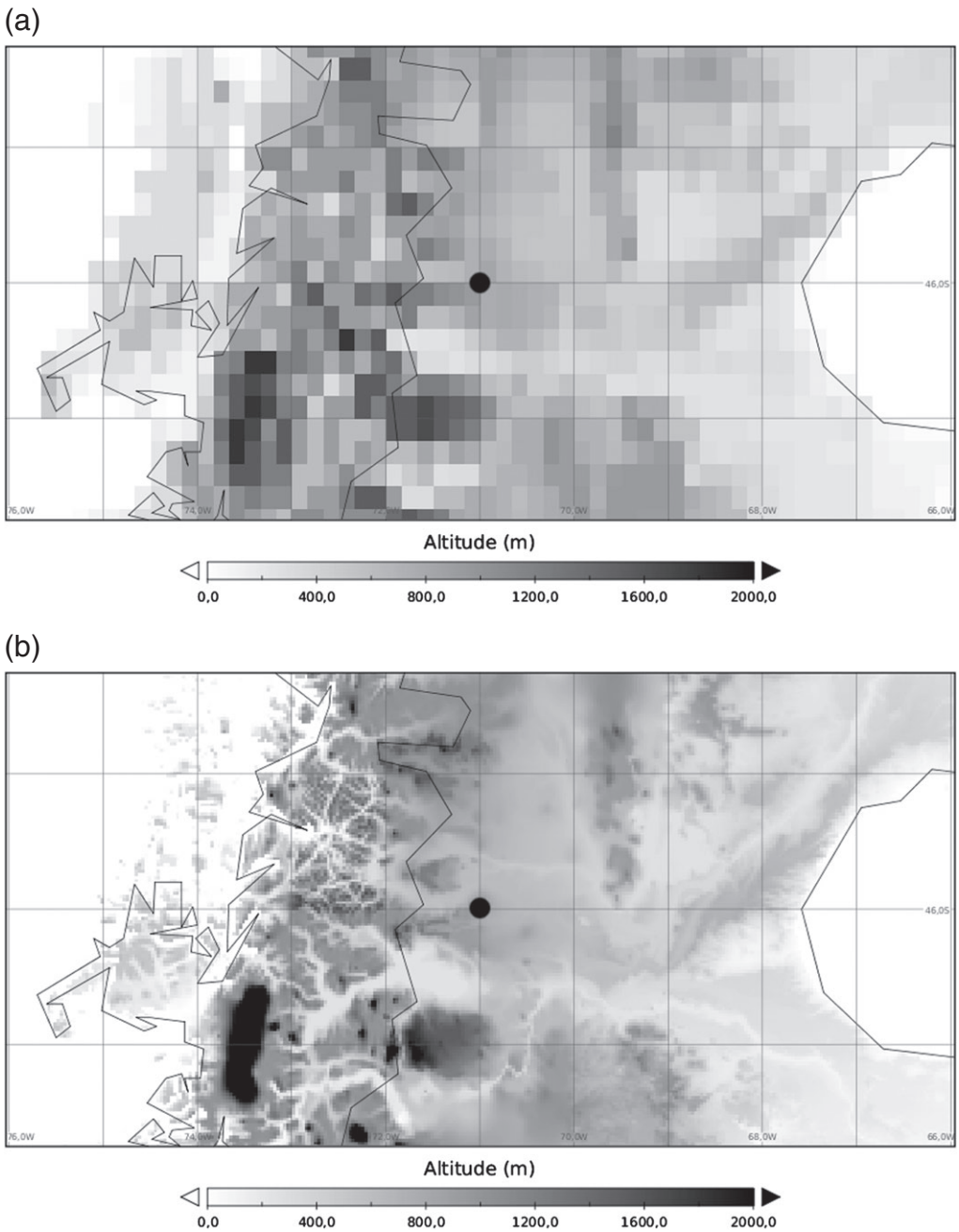

Figure 10. (a) Low $10' \times 10'$ and (b) high $2' \times 2'$ topographical resolution near location 46°S, 71°W (black dot) in the southern Andes.

estimation under the presence of model error. From a physical perspective, this bimodal result is not surprising since large H_{NC} yields low blocking levels, and a more efficient mountain elevation to excite gravity waves. When there is a larger amount of gravity wave drag, a good fraction of the corresponding gravity waves is likely to break at low level; this may be an effect hidden in the low-level drag discussed earlier. Also, this bimodality may be inherent in the nonlinear low-level flow dynamics the scheme tries to represent.

Figure 12 shows the five profiles of the SSO tendency intensity (i.e. the norm of the SSO tendency) generated with the estimated parameters for weak (5 July 2000) and strong (25 July 2000) surface wind conditions. In both surface wind conditions, the sets of estimated parameters with large G tend to underestimate the low-level drag (between 900 and 1000 hPa) and to overestimate the drag at high levels (between 650 and 900 hPa). On the other hand, the sets of estimated parameters with small G (and large H_{NC}) tend to overestimate the drag at low levels and also at higher levels (but they are relatively closer to the observed than in the cases with large G at those levels). The spread in the β parameter also appears to play a role.

6. Conclusion and outlook

In this article, we use a filtering technique to estimate the physical parameters of a subgrid-scale orographic scheme. The estimation is conducted offline, without estimating the state of the atmosphere and thus reducing the size of the state vector. As forcing terms, we use simulations of a general circulation model. The estimation problem is written as a nonlinear state-space system. This formulation is flexible and overcomes the main difficulties such as the boundaries on the physical parameters (strictly positive), the unknown background covariances and the high nonlinearity of the orographic scheme. In this state-space model, we suppose that the state and observation equations have additive Gaussian noise and that we know the *a priori* distribution of the physical parameters. The choice of these statistical parameters constitutes an important condition of convergence of the system to the true physical parameters. Thus, we estimate them via a maximum likelihood method. We use an iterative algorithm that computes the expected total log-likelihood function and maximizes it with respect to the statistical parameters.

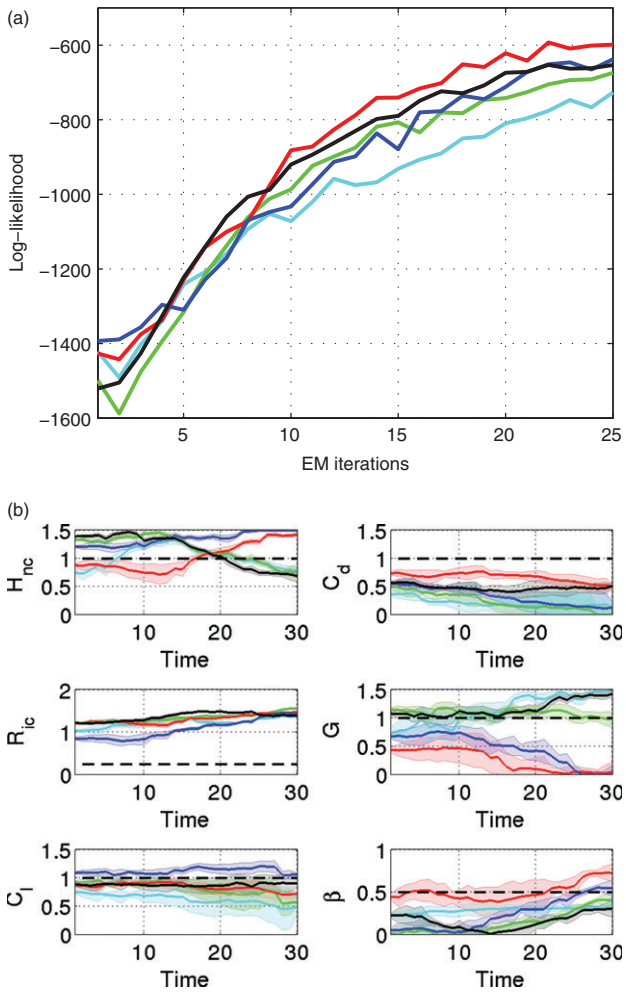

Figure 11. Evolution of five cases, with $N = 100$ members and the high-resolution orographic scheme, of (a) the innovation log-likelihood along the EM iterations and (b) the physical parameters estimated by the EnKF over time at iteration $j = 25$ of the EM algorithm. The straight dashed lines in (b) denote the physical parameter values of the low-resolution orographic scheme. In (b), the physical parameters ($\boldsymbol{\theta}$, not $\mathbf{x}$) are shown.

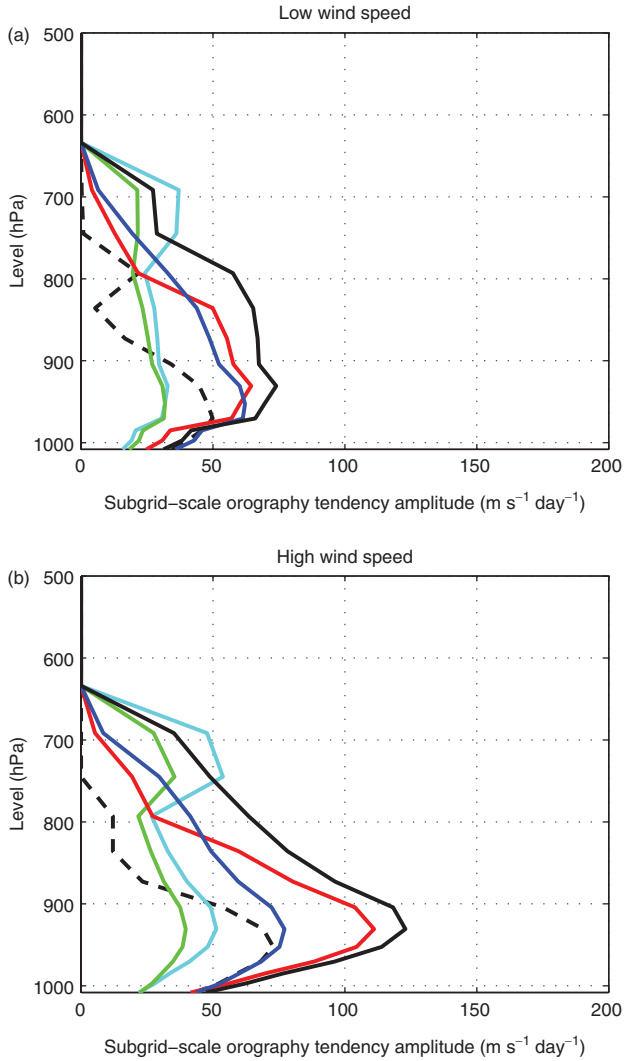

Figure 12. Profiles of the SSO tendency amplitude generated with the five sets of estimated parameters obtained with the SSO scheme using the high-resolution dataset on (a) 5 July 2000 and (b) 25 July 2000. The profiles used as observations and generated with the SSO scheme using the low-resolution dataset are shown by dashed lines.

The estimation technique is evaluated in a single vertical column and using synthetic observations (i.e. without using real observations but those produced by the SSO scheme). We imagine that an observational campaign takes place near the Perito Moreno Glacier in the Andes, where the topographical conditions are ideal to study mountain drag. First, we use twin experiments: we prescribe a true set of physical parameters and generate synthetic observations of mountain drag. Then, we apply the estimation technique using these generated observations and compare the estimated parameters to the true ones. The results indicate a convergence of the filter to the true parameters after $\sim$ 20 iterations of the EM algorithm. Even if the user initializes the error covariances and initial guess conditions with inappropriate values, these statistical parameters are iteratively updated and will converge towards the optimal values. The technique is able to detect correlations between parameters, to weight the observations as a function of the external forcing terms and to generate adaptive *a priori* information on the parameters. This overcomes the results obtained with deterministic values of statistical parameters which are usually arbitrarily prescribed since they are unknown.

We also examined whether the estimation technique is useful to determine whether the physical parameter should be changed when the horizontal resolution of an input dataset of the general circulation model is increased. In this case, the SSO scheme is imperfect and our filter takes into account this model error adding Gaussian noises controlled by time-dependent covariance matrices. The results show that our technique is a useful tool to determine the changes in the parameter when the resolution of the input orography dataset increases. However, model error degrades the estimated drag profiles; some features of the observed drag profile in Figure 12 cannot be reproduced by the estimated drag profiles that use the high-resolution orography dataset. A technique with model bias treatment as in Dee and Da Silva (1998) may be required to diminish the differences in the drag profiles. We also detected that some parameters may have a range of values for which the RMSE and the likelihood (cf. Figure 11(a)) almost do not change. These results show that there is no sensitivity to these parameters and therefore show that a precise value for these parameters is not important. We attribute this to the fact that, in the SSO scheme, a lot of drag is applied at low level and handled implicitly. In the scheme also, the low-level drag is multiplied by the number of ridges present in the gridbox area, a number which is around 1 or 2 for the US Navy $10' \times 10'$ dataset, and which becomes much larger when a more refined dataset is used. Ideally, we should take into account that, when a mountain exerts a drag, a wake downstream is associated with it, so that for mountains in the lee but still in the gridbox, the incident flow should be much reduced. Currently, when we increase the orography resolution, this effect is handled by an implicit treatment. Numerically this situation is satisfying, but clearly call for further understanding of the dynamical sheltering, and its impact on the large-scale flow. It may explain the difficulty in estimating the parameters and the difference between the drags generated with low- and high-resolution orographic datasets in Figure 12.

The technique presented here is an efficient method to resolve offline physical parameter estimation. The advantages are (i) the flexibility of the state-space formulation which can be applied to a large number of applications, (ii) the ability to estimate the background state and the eventually flow-dependent (and not necessarily diagonal) error covariance matrices **Q** and **R** of the EnKF, and (iii) the relative low computational cost of the technique where a relatively small number of members and few iterations of the EM algorithm are needed. One possible extension of the technique is the estimation of biases in more realistic cases with different kinds of model error.

This work is focused on the evaluation of the technique using first twin experiments and also an experiment with synthetic observations but using a higher-resolution orographic dataset so that the model (used in the data assimilation system) in this case is imperfect. In a real application, the technique requires vertical profiles of small-scale momentum forcing. We envisage two possible sources of this forcing which can be used to constrain orographic parameters. The most significant one is from intensive observational campaigns over mountainous areas. One of the most representative ones was PYREX (Bougeault *et al.*, 1990). Currently, there are several proposed campaigns (over the Andes, over New Zealand and over Scandinavia) for intense measurements over mountains with aircraft, lidars and radiosondes. These combined instruments can give significant information on momentum fluxes and their divergences. These potential campaigns could be an important source of observational data to estimate parameters of the subgrid-orography schemes using the proposed technique. A second possible data source of small-scale momentum forcing can be obtained from data assimilation techniques. Pulido and Thuburn (2005) show that four-dimensional variational assimilation can be used to estimate the missing momentum forcing term in the model equations. The technique is applied to obtain missing momentum forcing profiles in the middle atmosphere; here a significant part of systematic model error can be associated with gravity wave drag since the other physical parametrization active at those levels–the radiative transfer scheme–contains well-known parameters. On the other hand, in the troposphere several parametrizations are coupled so that the source of missing momentum is not readily identifiable with a particular parametrization. Therefore, the data assimilation techniques might be potentially useful to constrain subgrid-orography schemes using only the momentum forcing profile in the stratosphere. However, the impact of model errors from different sources in the parameter estimation problem needs to be further investigated. Another point that needs to be further investigated in an actual application of this offline technique is the possible feedbacks between the parametrization and the low-level flow; these feedback processes can affect the optimal parameters.

Follow-up work could apply this technique for online parameter estimation in strongly nonlinear systems. A first step will be to evaluate the method in a low-dimension system. Parameter estimation in a low-dimensional model was previously done by Annan and Hargreaves (2004) using deterministic values of the background state and the error covariance matrices. The advantage of applying our technique is to estimate them properly via the EM algorithm. Some first simulations we have performed give promising results. A simplified version of the method may also be useful in a larger-dimension online parameter estimation problem, for instance when there are a few unknown statistical parameters which need to be estimated precisely.

Acknowledgements

This work was supported by ANPCyT Argentina under grant PICT 2007 No. 411, the ECOS-Sud project DIAGAC and by the LEFE project IAC. The authors would like to thank Dr P. Ailliot and Dr J.J. Ruiz for their helpful comments.

Appendix A

E step

At each iteration j of the EM algorithm, we need the following conditional expectations:

$$
\begin{aligned}
&E\left[\mathbf{x}(t_k)\mathbf{x}(t_k)^\top|\mathbf{y}_{1:K},\widehat{\boldsymbol{\psi}}^{(j-1)}\right]\\
&\qquad=\mathbf{x}^{\mathrm{s}}(t_k)\mathbf{x}^{\mathrm{s}}(t_k)^\top+\mathbf{P}^{\mathrm{s}}(t_k)\\
&E\left[\mathbf{x}(t_k)\mathbf{x}(t_{k-1})^\top|\mathbf{y}_{1:K},\widehat{\boldsymbol{\psi}}^{(j-1)}\right]\\
&\qquad=\mathbf{x}^{\mathrm{s}}(t_k)\mathbf{x}^{\mathrm{s}}(t_{k-1})^\top+\mathbf{P}^{\mathrm{s}}(t_k,t_{k-1})\\
&E\left[\mathcal{H}_k\left\{\mathbf{x}(t_k)\right\}|\mathbf{y}_{1:K},\widehat{\boldsymbol{\psi}}^{(j-1)}\right]\\
&\qquad=\mathcal{H}_k\left\{\mathbf{x}^{\mathrm{s}}(t_k)\right\}\\
&E\left[\mathcal{H}_k\left\{\mathbf{x}(t_k)\right\}\mathcal{H}_k\left\{\mathbf{x}(t_k)\right\}^\top|\mathbf{y}_{1:K},\widehat{\boldsymbol{\psi}}^{(j-1)}\right]\\
&\qquad=\mathcal{H}_k\left\{\mathbf{x}^{\mathrm{s}}(t_k)\right\}\mathcal{H}_k\left\{\mathbf{x}^{\mathrm{s}}(t_k)\right\}^\top+\mathbf{P}^{\mathrm{s}}_{yy}(t_k),
\end{aligned}
$$

where $\mathbf{P}^{\mathrm{s}}_{yy}(t_k)$ is the sample covariance of the $\mathcal{H}_k\left\{\mathbf{x}^{\mathrm{s}}_i(t_k)\right\}\forall i\in\{1,\ldots,N\}$.

Appendix B

M step

The maximum likelihood estimates of the statistical parameters are given by:

$$
\begin{aligned}
\widehat{\mathbf{x}}^{(j)}_{\mathrm{b}}&=E\left[\mathbf{x}(t_1)|\mathbf{y}_{1:K},\widehat{\boldsymbol{\psi}}^{(j-1)}\right],\\
\widehat{\mathbf{B}}^{(j)}&=\mathrm{Var}\left[\mathbf{x}(t_1)|\mathbf{y}_{1:K},\widehat{\boldsymbol{\psi}}^{(j-1)}\right],\\
\widehat{\mathbf{Q}}^{(j)}&=\frac{1}{T-1}\sum_{k=2}^{T}E\left[\mathbf{x}(t_k)\mathbf{x}(t_k)^\top|\mathbf{y}_{1:K},\widehat{\boldsymbol{\psi}}^{(j-1)}\right]\\
&\quad-\frac{1}{T-1}\sum_{k=2}^{T}E\left[\mathbf{x}(t_k)\mathbf{x}(t_{k-1})^\top|\mathbf{y}_{1:K},\widehat{\boldsymbol{\psi}}^{(j-1)}\right]\\
&\quad-\frac{1}{T-1}\sum_{k=2}^{T}E\left[\mathbf{x}(t_k)\mathbf{x}(t_{k-1})^\top|\mathbf{y}_{1:K},\widehat{\boldsymbol{\psi}}^{(j-1)}\right]^\top\\
&\quad+\frac{1}{T-1}\sum_{k=2}^{T}E\left[\mathbf{x}(t_{k-1})\mathbf{x}(t_{k-1})^\top|\mathbf{y}_{1:K},\widehat{\boldsymbol{\psi}}^{(j-1)}\right],\\
\widehat{\mathbf{R}}^{(j)}(t_k)&=\mathbf{y}(t_k)\mathbf{y}(t_k)^\top\\
&\quad-E\left[\mathcal{H}_k\left\{\mathbf{x}(t_k)\right\}|\mathbf{y}_{1:K},\widehat{\boldsymbol{\psi}}^{(j-1)}\right]\mathbf{y}(t_k)^\top\\
&\quad-\mathbf{y}(t_k)E\left[\mathcal{H}_k\left\{\mathbf{x}(t_k)\right\}|\mathbf{y}_{1:K},\widehat{\boldsymbol{\psi}}^{(j-1)}\right]^\top\\
&\quad+E\left[\mathcal{H}_k\left\{\mathbf{x}(t_k)\right\}\mathcal{H}_k\left\{\mathbf{x}(t_k)\right\}^\top|\,\mathbf{y}_{1:K},\widehat{\boldsymbol{\psi}}^{(j-1)}\right],
\end{aligned}
$$

where the conditional expectations are computed in the E step via the EnKS.

Appendix C

Innovation likelihood

The innovation likelihood function is given by

$$
\begin{aligned}
l(\mathbf{x},\boldsymbol{\psi})=&\\
&\prod_{k=1}^{K}\exp\left[-\frac{1}{2}\mathbf{d}(t_k)^\top\left\{\mathbf{P}^{\mathrm{f}}_{yy}(t_k)+\mathbf{R}(t_k)\right\}^{-1}\mathbf{d}(t_k)\right]\\
&\qquad\times(2\pi)^{-p/2}\left[\det\left\{\mathbf{P}^{\mathrm{f}}_{yy}(t_k)+\mathbf{R}(t_k)\right\}\right]^{-1/2},
\end{aligned}
$$

with the covariance matrix $\mathbf{P}^{\mathrm{f}}_{yy}(t_k)$ and the innovation vector $\mathbf{d}(t_k)$ given in Eqs (12) and (13) respectively.

References

Anderson JL. 2007. An adaptive covariance inflation error correction algorithm for ensemble filters. *Tellus* **59A**: 210–224.

Annan JD, Hargreaves JC. 2004. Efficient parameter estimation for a highly chaotic system. *Tellus* **56A**: 520–526.

Annan JD, Hargreaves JC. 2007. Efficient estimation and ensemble generation in climate modelling. *Phil. Trans. R. Soc. A: Math. Phys. Eng. Sci.* **365**: 2077–2088.

Bougeault P, Jansa Clar A, Benech B, Carissimo B, Pelon J, Richard E. 1990. Momentum budget over the Pyrénées: The PYREX experiment. *Bull. Am. Meteorol. Soc.* **71**: 806–818.

Burgers G, Van Leeuwen PJ, Evensen G. 1998. Analysis scheme in the ensemble Kalman filter. *Mon. Weather Rev.* **126**: 1719–1724.

Cappé O, Moulines E, RydØn T. 2005. *Inference in Hidden Markov Models*. Springer Science+Business Media: New York, NY.

Dee DP, Da Silva AM. 1998. Data assimilation in the presence of forecast bias. *Q. J. R. Meteorol. Soc.* **124**: 269–295.

Dempster AP, Laird NM, Rubin DB. 1977. Maximum likelihood from incomplete data via the EM algorithm. *J. R. Stat. Soc. Ser. B* **39**: 1–38.

Evensen G. 2009. *Data Assimilation: The Ensemble Kalman Filter*. Springer-Verlag: Berlin.

Evensen G, Van Leeuwen PJ. 2000. An ensemble Kalman smoother for nonlinear dynamics. *Mon. Weather Rev.* **128**: 1852–1867.

Hertzog A, Alexander MJ, Plougonven R. 2012. On the intermittency of gravity wave momentum flux in the stratosphere. *J. Atmos. Sci.* **69**: 3433–3448.

Hourdin F, Musat I, Bony S, Braconnot P, Codron F, Dufresne J-L, Fairhead L, Filiberti M-A, Friedlingstein P, Grandpeix J-Y, Krinner G, Levan P, Li Z-X, Lott F. 2006. The LMDZ4 general circulation model: Climate performance and sensitivity to parametrized physics with emphasis on tropical convection. *Clim. Dyn.* **27**: 787–813.

Hu XM, Zhang F, Nielsen-Gammon JW. 2010. Ensemble-based simultaneous state and parameter estimation for treatment of mesoscale model error: A real-data study. *Geophys. Res. Lett.* **37**: L08802, doi: 10.1029/2010GL043017.

Ide K, Courtier P, Ghil M, Lorenc AC. 1997. Unified notation for data assimilation: Operational, sequential and variational. *J. Meteorol. Soc. Jpn.* **75**: 181–189.

Jackson C, Sen MK, Stoffa PL. 2004. An efficient stochastic Bayesian approach to optimal parameter and uncertainty estimation for climate model predictions. *J. Clim.* **17**: 2828–2841.

Li F, Austin J, Wilson J. 2008. The strength of the Brewer–Dobson circulation in a changing climate: Coupled chemistry–climate model simulations. *J. Clim.* **21**: 40–57.

Li H, Kalnay E, Miyoshi T. 2009. Simultaneous estimation of covariance inflation and observation errors within an ensemble Kalman filter. *Q. J. R. Meteorol. Soc.* **135**: 523–533.

Liang X, Zheng X, Zhang S, Wu G, Dai Y, Li Y. 2011. Maximum likelihood estimation of inflation factors on error covariance matrices for ensemble Kalman filter assimilation. *Q. J. R. Meteorol. Soc.* **138**: 263–273.

Lott F. 1995. Comparison between the orographic response of the ECMWF model and the PYREX 1990 data. *Q. J. R. Meteorol. Soc.* **121**: 1323–1348.

Lott F. 1999. Alleviation of stationary biases in a GCM through a mountain drag parameterization scheme and a simple representation of mountain lift forces. *Mon. Weather Rev.* **127**: 788–801.

Lott F, Miller MJ. 1997. A new subgrid-scale orographic drag parametrization: Its formulation and testing. *Q. J. R. Meteorol. Soc.* **123**: 101–127.

Lott F, Fairhead L, Hourdin F, Levan P. 2005. The stratospheric version of LMDz: Dynamical climatologies, Arctic oscillation, and impact on the surface climate. *Clim. Dyn.* **25**: 851–868.

McLandress C, Shepherd TG. 2009. Simulated anthropogenic changes in the Brewer–Dobson circulation, including its extension to high latitudes. *J. Clim.* **22**: 1513–1540.

Miyoshi T. 2011. The Gaussian approach to adaptive covariance inflation and its implementation with the local ensemble transform Kalman filter. *Mon. Weather Rev.* **139**: 1519–1535.

Miyoshi T, Kalnay E, Li H. 2012. Estimating and including observation-error correlations in data assimilation. *Inverse Prob. Sci. Eng.* **21**: 387–398, doi: 10.1080/17415977.2012.712527.

NOAA. 2001. *NOAA ETOPO2 Dataset, 2-minute Gridded Global Relief Data*. National Geophysical Data Center: Boulder, CO. http://www.ngdc.noaa.gov/mgg/fliers/01mgg04.html (accessed 13 March 2014).

Palmer TN, Shutts GJ, Swinbank R. 1986. Alleviation of a systematic westerly bias in general circulation and numerical weather prediction models through an orographic gravity wave drag parametrization. *Q. J. R. Meteorol. Soc.* **112**: 1001–1039.

Pham DT. 2001. Stochastic methods for sequential data assimilation in strongly nonlinear systems. *Mon. Weather Rev.* **129**: 1194–1207.

Posselt DJ, Bishop CH. 2012. Nonlinear parameter estimation: Comparison of an ensemble Kalman smoother with a Markov chain Monte Carlo algorithm. *Mon. Weather Rev.* **140**: 1957–1974.

Pulido M, Thuburn J. 2005. Gravity wave drag estimation from global analyses using variational data assimilation principles. I: Theory and implementation. *Q. J. R. Meteorol. Soc.* **131**: 1821–1840.

Pulido M, Thuburn J. 2008. The seasonal cycle of gravity wave drag in the middle atmosphere. *J. Clim.* **21**: 4664–4679.

Pulido M, Polavarapu S, Shepherd TG, Thuburn J. 2012. Estimation of optimal gravity wave parameters for climate models using data assimilation. *Q. J. R. Meteorol. Soc.* **138**: 298–309.

Ruiz JJ, Pulido M, Miyoshi T. 2013. Estimating model parameters with ensemble-based data assimilation: A review. *J. Meteorol. Soc. Jpn.* **91**: 79–99.

Schirber S, Klocke D, Pincus R, Quaas J, Anderson JL. 2013. Parameter estimation using data assimilation in an atmospheric general circulation model: From a perfect toward the real world. *J. Adv. Model. Earth Syst.* **5**: 58–70.

Severijns CA, Hazeleger W. 2005. Optimizing parameters in an atmospheric general circulation model. *J. Clim.* **18**: 3527–3535.

Sigmond M, Scinocca JF, Kushner PJ. 2008. Impact of the stratosphere on tropospheric climate change. *Geophys. Res. Lett.* **35**: L12706, doi: 10.1029/2008GL033573.

Stainforth DA, Aina T, Christensen C, Collins M, Faull N, Frame DJ, Kettleborough JA, Knight S, Martin A, Murphy JM, Piani C, Sexton D, Smith LA, Spicer RA, Thorpe AJ, Allen MR. 2005. Uncertainty in predictions of the climate response to rising levels of greenhouse gases. *Nature* **433**: 403–406.

Tandeo P, Ailliot P, Autret E. 2011. Linear Gaussian state-space model with irregular sampling: Application to sea surface temperature. *Stoch. Environ. Res. Risk Assess.* **25**: 793–804.

Wang X, Bishop CH. 2003. A comparison of breeding and ensemble transform Kalman filter ensemble forecast schemes. *J. Atmos. Sci.* **60**: 1140–1158.

Wu G, Zheng X, Wang L, Zhang S, Liang X, Li Y. 2013. A new structure for error covariance matrices and their adaptive estimation in EnKF assimilation. *Q. J. R. Meteorol. Soc.* **139**: 795–804, doi: 10.1002/qj.2000.

Yang X, Delsole T. 2009. Using the ensemble Kalman filter to estimate multiplicative model parameters. *Tellus* **61A**: 601–609.

