An approach to estimating and extrapolating model error based on inverse problem methods: towards accurate numerical weather prediction*

Hu Shu-Juan(胡淑娟)[a)†], Qiu Chun-Yu(邱春雨)[a)b)], Zhang Li-Yun(张利云)[b)], Huang Qi-Can(黄启灿)[a)b)], Yu Hai-Peng(于海鹏)[a)], and Chou Ji-Fan(丑纪范)[a)]

[a)] *College of Atmospheric Sciences, Lanzhou University, Lanzhou* 730000, *China*
[b)] *School of Mathematics and Statistics, Lanzhou University, Lanzhou* 730000, *China*

(Received 17 December 2013; revised manuscript received 29 January 2014; published online 10 June 2014)

Model error is one of the key factors restricting the accuracy of numerical weather prediction (NWP). Considering the continuous evolution of the atmosphere, the observed data (ignoring the measurement error) can be viewed as a series of solutions of an accurate model governing the actual atmosphere. Model error is represented as an unknown term in the accurate model, thus NWP can be considered as an inverse problem to uncover the unknown error term. The inverse problem models can absorb long periods of observed data to generate model error correction procedures. They thus resolve the deficiency and faultiness of the NWP schemes employing only the initial-time data. In this study we construct two inverse problem models to estimate and extrapolate the time-varying and spatial-varying model errors in both the historical and forecast periods by using recent observations and analogue phenomena of the atmosphere. Numerical experiment on Burgers' equation has illustrated the substantial forecast improvement using inverse problem algorithms. The proposed inverse problem methods of suppressing NWP errors will be useful in future high accuracy applications of NWP.

Keywords: numerical weather prediction, model error, past data, inverse problem

PACS: 92.60.Wc **DOI:** 10.1088/1674-1056/23/8/089201

1. Introduction

No matter how well we understand the physical processes of the actual atmosphere, numerical weather prediction (NWP) error originates from two distinct sources: inaccurate initial conditions and model errors. Many studies have focused on the errors of initial conditions including data assimilation and ensemble forecasting techniques.[1–7] More accurate estimation of the initial conditions is necessary, but it is not sufficient for improving weather forecasts.[8,9] Model errors, which include deficiencies in numerical differentiation and integration as well as inaccurate forcing and parameterizations of the subgrid-scale physical processes, result in large-scale systematic errors of NWP. Currently, many more efforts to tackle model errors have focused on generating more accurate approximations of forcing, improving the parameterizations of physical processes, and increasing the grid density to resolve smaller-scale processes. However, unresolved phenomena and model errors will be present no matter how precise the forcing is, how accurate the parameterizations are, and how fine the grid resolution is. Therefore, it is important to develop mathematical methods to account for model errors.[10]

Leith[11] proposed a statistical method to diagnose model bias and systematic errors linearly dependent on state anomalies. However, Leith's empirically estimated state-dependent correction term is optimal only for a linear model. Inspired by Leith's procedure, various approaches to model error correction have been presented and produced promising results. For example, the singular value decomposition (SVD) method[10,12,13] and empirical orthogonal function (EOF) technique [14] are used to estimate the state-independent (constant) or state-dependent model error (function of model state). These studies are based on the traditional viewpoint that NWP is in essence an initial-value problem (namely, forward problem). Gu[15] pointed out that the fundamental defect of the NWP is only using single-time data in the initial-value problem. Chou[16] proposed a theoretical method to improve the NWP by using past data, in which the observed data (ignoring the measurement error) are considered as a series of solutions of the accurate model and then the NWP with an unknown error term actually becomes an inverse problem. The unknown model error term should be determined by the solutions (observed data) of the partial differential equation of the actual atmospheric motion. If we can solve the inverse problem effectively, then, by adding an error term (the solution of the inverse problem) to the model, we can inevitably improve the NWP. In contrast to the existing model error correction methods based on forward problems, the inverse prob-

*Project supported by the Special Scientific Research Project for Public Interest (Grant No. GYHY201206009), the Fundamental Research Funds for the Central Universities, China (Grant Nos. lzujbky-2012-13 and lzujbky-2013-11), and the National Basic Research Program of China (Grant Nos. 2012CB955902 and 2013CB430204).

†Corresponding author. E-mail: hushuju@lzu.edu.cn

© 2014 Chinese Physical Society and IOP Publishing Ltd

lem method can internally use past data and improve forecast both dynamically and statistically.

A series of methods to estimate the forecast errors or model errors has been developed by using the inverse problem ideas. For example, Qiu and Chou[17–19] presented a pattern recognition and parameter optimizing method. Huang and Yi[20] and Huang *et al.*[21] established a coupled analogue deviation model for monthly and seasonal forecasts and achieved results better than those from the model for a statistical analogue forecast. Based on this method, Bao *et al.*,[22] Ren *et al.*,[23] and Zheng *et al.*[24] estimated forecast error instead of tendency error by using analogue techniques which avoid establishing an analogue deviation model. They used this method in an operational T63L16 model and obtained considerable predictive skill in monthly means and extended range forecasts. Meanwhile, with the refinement and increase of the observational network, much research has been devoted to estimating and accounting for model error in or using variational data assimilation methods and is capable of extracting significant information about the model error from the observations.[25–27] However, because of the uncertainty of the factors causing model errors, these studies usually decompose the model errors into systematic bias, a model state-dependent part and a random error part, and so on,[28] where the covariances and/or dynamics of model error are often needed. An accurate estimate of the observation error covariance is usually at hand, while more difficulties arise from the background and model error covariances.[27] At the same time, due to the lack of the knowledge of the accurate dynamics of model error, a simple dynamical law, such as a first-order Markov process, has usually been used.[25] For these defects, a new issue in the estimation of model error as a missing forcing rather than its empirical decomposed parts is presented.[29–31]

The purpose of the present study is to explore the feasibility of estimating and extrapolating the time-varying and spatial-varying model errors using inverse problem methods and past data, which objectively suppose the model error as an unknown functional term (a missing forcing term) of the numerical model. The rest of this paper is organized as follows. In Section 2, we construct an inverse problem to compute the time-varying and spatial-varying model error dataset by using the past observations, in which the observed data are actually viewed as the dynamics of model error. In Section 3, we construct another inverse problem to extrapolate the time-varying and spatial-varying model errors in the forecast period according to the historical model error dataset. The numerical experiments are presented in Section 4 to illustrate the substantial forecast improvement resulting from our inverse problem algorithms. The main results are summarized and discussed in the concluding section, in which we point out that our inverse problem methods can be used especially in local NWP models to improve the prediction of the local investigated phenomena.

2. Inverse problem for model error

2.1. Inverse problem

Without loss of generality, if we do not consider the effect of boundary conditions, the NWP model can be written as the initial problem

$$\begin{cases} \dfrac{\partial \psi(\boldsymbol{x},t)}{\partial t} + L(\psi(\boldsymbol{x},t)) = 0, \quad \boldsymbol{x} \in D, t > t_0, \\ \psi|_{t=t_0} = \psi_0(\boldsymbol{x}), \end{cases} \tag{1}$$

where D denotes the range of spatial variables x, ψ and ψ_0 are the state vector (including wind, temperature, moisture, pressure, etc.) and initial condition, respectively, and L is the nonlinear differential operator governing the dynamics of the model system. Because of the deficiencies in parameterizations of physical processes and inaccurate forcing, equation (1) is usually an approximation of the actual atmosphere. If we use A to represent the nonlinear differential operator governing the dynamics of the actual atmosphere, the accurate model (the ultimate aim of all numerical models) of the true system can be written as

$$\begin{cases} \dfrac{\partial \psi(\boldsymbol{x},t)}{\partial t} + A(\psi(\boldsymbol{x},t)) = 0, \quad \boldsymbol{x} \in D, t > t_0, \\ \psi|_{t=t_0} = \psi_0(\boldsymbol{x}), \end{cases} \tag{2}$$

where A is the nonlinear differential operator governing the dynamics of the actual atmosphere.

Let $E(\psi(\boldsymbol{x},t)) = A(\psi(\boldsymbol{x},t)) - L(\psi(\boldsymbol{x},t))$, then equation (2) can be rewritten as

$$\begin{cases} \dfrac{\partial \psi(\boldsymbol{x},t)}{\partial t} + L(\psi(\boldsymbol{x},t)) + E(\psi(\boldsymbol{x},t)) = 0, \quad \boldsymbol{x} \in D, t > t_0, \\ \psi|_{t=t_0} = \psi_0(\boldsymbol{x}), \end{cases} \tag{3}$$

where E represents the unknown model error term (the missing forcing term) that is not or inaccurately described by Eq. (1). The model error term $E(\psi(\boldsymbol{x},t))$ is both time varying and spatial varying.

From the viewpoint of dynamics, the existing observed data represent the true dynamic system of the actual atmosphere. The observed data (ignoring the measurement error) are actually a series of solutions of Eq. (3). Namely, we have known not only the initial value $\psi|_{t=t_0} = \psi_0(\boldsymbol{x})$ but also the solutions $\psi|_{t=t_1} = \psi_1(\boldsymbol{x}), \ldots, \psi|_{t=t_K} = \psi_K(\boldsymbol{x})$ of Eq. (3). The information about the unknown model error term $E(\psi(\boldsymbol{x},t))$ is included in the past data $\psi_0, \psi_1, \ldots, \psi_K$. Problem (3) with the unknown term $E(\psi(\boldsymbol{x},t))$ is a standard partial differential equation inverse problem.

Unlike forward problems, system (3) with solutions of $\psi_0, \psi_1, \ldots, \psi_K$ is usually overdetermined and ill posed. The

common method to deal with this problem is to seek the optimal fitting of the minimization problem

$$\min_{E} \frac{1}{2} \iiint_{D} \sum_{k=1}^{K} |\psi(\boldsymbol{x},t_k) - \psi_k(\boldsymbol{x})|^2 \,\mathrm{d}\boldsymbol{x}, \tag{4}$$

where $\psi(\boldsymbol{x},t_k)$ and $\psi_k(\boldsymbol{x})$ are the solution of initial value problem (3) and observed data at time t_k, respectively. Minimization problem (4) constrained by Eq. (3) means that we try to find the error term $E(\psi(\boldsymbol{x},t))$ such that the forecast of Eq. (3) is as close to the observation as possible. However, problem (4) is still ill posed because it has a non-unique solution, which will lead to instability of the numerical computation. As a consequence of the regularization theory for inverse problems,[32] we should penalize the problem (4) and construct the following optimization problem:

$$\begin{aligned}\min_{E} &\frac{1}{2} \iiint_{D} \sum_{k=1}^{K} |\psi(\boldsymbol{x},t_k) - \psi_k(\boldsymbol{x})|^2 \,\mathrm{d}\boldsymbol{x} \\ &+ \frac{\alpha}{2} \int_{t_0}^{t_K} \iiint_{D} |E(\psi(\boldsymbol{x},t))|^2 \,\mathrm{d}\boldsymbol{x}\,\mathrm{d}t,\end{aligned} \tag{5}$$

where $\alpha > 0$ is the regularization parameter. The second term, $E(\psi(\boldsymbol{x},t))$, of expression (5) indicates that the unknown model error term $E(\psi(\boldsymbol{x},t))$ is actually a small quantity in Eq. (3), which admits the reliability of numerical model (1) and, at the same time, resolves the ill-posedness of problem (4).

2.2. Discrete problem

The solution $E(\psi(\boldsymbol{x},t))$ of optimization problem (5) constrained by Eq. (3) is just the model error of numerical model (1) in time interval $t_0 \le t \le t_K$. Unfortunately, we cannot obtain the minimum solution by solving the infinite dimensional problem, but instead consider its discretization. First, to discretize expression (5) and Eq. (3) in space, we use the Galerkin method. If $\varphi_0(\boldsymbol{x}), \varphi_1(\boldsymbol{x}), \ldots, \varphi_n(\boldsymbol{x})$ are a series of basis functions that satisfy the boundary conditions of problem (3), then we approximate $\psi(\boldsymbol{x},t)$ and $E(\psi(\boldsymbol{x},t))$ by functions of the form

$$\psi_h(\boldsymbol{x},t) = \sum_{j=0}^{n} \alpha_j(t)\varphi_j(\boldsymbol{x}), \tag{6}$$

and

$$E_h(\psi(\boldsymbol{x},t)) = \sum_{j=0}^{n} e_j(t)\varphi_j(\boldsymbol{x}). \tag{7}$$

If we insert Eqs. (6) and (7) into the objective function of expression (5) and multiply the differential equation (3) by $\varphi_i(\boldsymbol{x})$ $(i = 0,1,\ldots,n)$ and integrate both sides over spatial variable domain D, we obtain

$$\begin{aligned}\min_{\boldsymbol{E}(t)} \sum_{k=1}^{K} \left(\frac{1}{2}\psi(t_k)^{\mathrm{T}} M \psi(t_k) + \boldsymbol{g}(t_k)^{\mathrm{T}} \psi(t_k) \right) \\ + \frac{\alpha}{2} \int_{t_0}^{t_K} \boldsymbol{E}(t)^{\mathrm{T}} M \boldsymbol{E}(t)\,\mathrm{d}t,\end{aligned} \tag{8}$$

where $\psi(t_k)$ is the solution of the following initial problem:

$$\begin{cases} M\dfrac{\mathrm{d}}{\mathrm{d}t}\psi(t) + N(\psi(t)) + M\boldsymbol{E}(t) = 0, \quad t_0 < t < t_K, \\ \psi(t_0) = \psi_0, \end{cases} \tag{9}$$

at time t_k, $\psi(t) = (\alpha_0(t), \alpha_1(t), \ldots, \alpha_n(t))^{\mathrm{T}}$, $\boldsymbol{E}(t) = (e_0(t), e_1(t), \ldots, e_n(t))^{\mathrm{T}}$. $M \in R^{(n+1)\times(n+1)}$ is a matrix, and $N(\psi(t)) \in R^{(n+1)}$ and $\boldsymbol{g}(t_k) \in R^{(n+1)}$ are vectors with elements

$$\begin{aligned}(M)_{ij} &= \iiint_{D} \varphi_i(\boldsymbol{x})\varphi_j(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x}, \\ N(\psi(t))_i &= \iiint_{D} L\left(\sum_{j=0}^{n} \alpha_j(t)\varphi_j(\boldsymbol{x}) \right) \varphi_i(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x}, \\ \boldsymbol{g}(t_k)_i &= -\iiint_{D} \psi_k(\boldsymbol{x})\varphi_i(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x}.\end{aligned}$$

To discretize expressions (8) and (9) in time, we use the Crank–Nicolson method. We let $t_0 < t_1 < \cdots < t_{mK} = t_K$ and define $\Delta t_i = t_{i+1} - t_i$ $(i = 0,\ldots,mK-1)$, where m is a given integer. We also introduce $\Delta t_{-1} = \Delta t_{mK} = 0$. Then the fully discretized forms of expressions (5) and (3) are given by

$$\begin{aligned}\min_{\boldsymbol{E}_0,\ldots,\boldsymbol{E}_{mK}} &\sum_{k=1}^{K} \left(\frac{1}{2}\psi_{mk}^{\mathrm{T}} M \psi_{mk} + \boldsymbol{g}_{mk}^{\mathrm{T}} \psi_{mk} \right) \\ &+ \frac{\alpha}{2} \sum_{i=0}^{mK} \frac{\Delta t_{i-1} + \Delta t_i}{2} \boldsymbol{E}_i^{\mathrm{T}} M \boldsymbol{E}_i,\end{aligned} \tag{10}$$

where $\boldsymbol{E}_i = \boldsymbol{E}_i(t_i)$ $(i = 0,1,\ldots,mK)$ and $\psi_{mk} = \psi(t_{mk})$ $(k = 1,\ldots,K)$ are the solutions of

$$\begin{aligned}M\psi_{i+1} + \frac{\Delta t_i}{2} N(\psi_{i+1}) + \frac{\Delta t_i}{2} M\boldsymbol{E}_{i+1} \\ - M\psi_i + \frac{\Delta t_i}{2} N(\psi_i) + \frac{\Delta t_i}{2} M\boldsymbol{E}_i = 0,\end{aligned} \tag{11}$$

where $i = 0,1,\ldots,mK-1$, and ψ_0 is given.

2.3. Algorithm

For convenience, we set $\boldsymbol{E} = (\boldsymbol{E}_0^{\mathrm{T}}, \boldsymbol{E}_1^{\mathrm{T}}, \ldots, \boldsymbol{E}_{mK}^{\mathrm{T}})^{\mathrm{T}}$, $\psi = (\psi_1^{\mathrm{T}}, \ldots, \psi_{mK}^{\mathrm{T}})^{\mathrm{T}}$, and we denote the objective function in expression (10) by $f(\psi,\boldsymbol{E})$ and in Eq. (11) by $c(\psi,\boldsymbol{E}) = 0$. Then we rewrite expression (10) and Eq. (11) as a constrained optimization problem

$$\begin{aligned}&\min_{\boldsymbol{E}} f(\psi,\boldsymbol{E}), \\ &\text{s.t.} \quad c(\psi,\boldsymbol{E}) = 0.\end{aligned} \tag{12}$$

Furthermore, if we insert the constrained equation $c(\psi,\boldsymbol{E}) = 0$ into the objective function $f(\psi,\boldsymbol{E})$ in Eq. (12), then we have an implicitly constrained optimization problem

$$\min_{\boldsymbol{E}} \hat{f}(\boldsymbol{E}), \tag{13}$$

where $\hat{f}(\boldsymbol{E}) = f(\boldsymbol{\psi}(\boldsymbol{E}), \boldsymbol{E})$, $\boldsymbol{\psi} = \boldsymbol{\psi}(\boldsymbol{E})$ is a functional of E determined by $c(\boldsymbol{\psi}, \boldsymbol{E}) = 0$. The numerical calculation of problem (13) using gradient-based and Newton-type methods is discussed in most courses on numerical analysis and in courses on optimization. However, the optimization of problem (13) on the surface looks relatively simple, but it is actually quite complicated because the gradient and Hessian computations for objective function $\hat{f}(\boldsymbol{E})$ involves the solution of the implicit nonlinear constraint $c(\boldsymbol{\psi}, \boldsymbol{E}) = 0$, namely the solution of the NWP model. We here state a commonly used version of the Newton–conjugate gradient (CG) method[33] for solving problem (13) in the following.

Algorithm 1 (Newton-CG method with Armijo line-search).

A-1. Given $\boldsymbol{E}_0$ and $gtol > 0$. Set $k = 0$.

A-2. Compute $\nabla \hat{f}(\boldsymbol{E}_k)$.

A-3. If $\left\|\nabla \hat{f}(\boldsymbol{E}_k)\right\| < gtol$ stop.

A-4. Compute $\nabla^2 \hat{f}(\boldsymbol{E}_k)$.

A-5. Apply the CG method to compute the Newton equation

$$\nabla^2 \hat{f}(\boldsymbol{E}_k) s_k = -\nabla \hat{f}(\boldsymbol{E}_k).$$

A-6. Perform Armijo line-search.

A-6.1. Set $h_k = 1$ and evaluate $\hat{f}(\boldsymbol{E}_k + h_k s_k)$.

A-6.2. While $\hat{f}(\boldsymbol{E}_k + h_k s_k) > \hat{f}(\boldsymbol{E}_k) + 10^{-4} h_k s_k^{\mathrm{T}} \nabla \hat{f}(\boldsymbol{E}_k)$ set $h_k = h_k/2$ and evaluate $\hat{f}(\boldsymbol{E}_k + h_k s_k)$.

A-7. Set $\boldsymbol{E}_{k+1} = \boldsymbol{E}_k + h_k s_k$, $k = k+1$. Goto A-2.

3. Inverse problem for extrapolation

3.1. Inverse problem

According to the optimization problem defined by expressions (5) and (3), we have obtained the time-varying and spatial-varying model error $E(\psi(\boldsymbol{x},t))$ in the historical period (in time interval $t_0 \le t \le t_K$). To improve the forecast, we now need to extrapolate the model error $E(\psi(\boldsymbol{x},t))$ to the forecast period. To simplify the expression, we use the notation $E(\boldsymbol{x},t)$ instead of $E(\psi(\boldsymbol{x},t))$ and use $\hat{E}(\boldsymbol{x},t)$ to denote the model error dataset in the historical period. We also use $O^{\mathrm{c}} = \{(\boldsymbol{x},t)|\boldsymbol{x} \in D, t \in [t_0, t_K]\}$ and $O = \{(\boldsymbol{x},t)|\boldsymbol{x} \in D, t \in [t_K, t_{K+1}]\}$ to represent the historical period (in the spatial domain D and time interval $t_0 \le t \le t_K$) and forecast period (in the spatial domain D and time interval $t_K < t \le t_{K+1}$), respectively. Then the extrapolation problem is to find model error $E(\boldsymbol{x},t)$ such that

$$E(\boldsymbol{x},t) = \begin{cases} E(\boldsymbol{x},t), & (\boldsymbol{x},t) \in O, \\ \hat{E}(\boldsymbol{x},t), & (\boldsymbol{x},t) \in O^{\mathrm{c}}. \end{cases} \tag{14}$$

Clearly the identity (14) can also be written as

$$\Phi(E(\boldsymbol{x},t)) = b(\boldsymbol{x},t) = \begin{cases} 0, & \text{if } (\boldsymbol{x},t) \in O, \\ \hat{E}(\boldsymbol{x},t), & \text{if } (\boldsymbol{x},t) \in O^{\mathrm{c}}, \end{cases} \tag{15}$$

where the matrix Φ is known as a projection operator from $O^{\mathrm{c}} \cup O$ to O^{c}. The right-hand side zero in Eq. (15) presents the first guess of the extrapolation value of model error in the forecast period, since no better guess is available. Equation (15) is another standard inverse problem, namely we need to compute $E(\boldsymbol{x},t)$ in both historical period and forecast period by using matrix Φ and right side $b(\boldsymbol{x},t)$ in Eq. (15). Then, according to the regularization theory for inverse problems,[32] we have the following optimization problem:

$$\min_{E|_{O^{\mathrm{c}} \cup O}} \frac{1}{2} \int_{O^{\mathrm{c}} \cup O} |\Phi(E(\boldsymbol{x},t)) - b(\boldsymbol{x},t)| \, \mathrm{d}\boldsymbol{x} \, \mathrm{d}t + \frac{\alpha}{2} R(E)|_{O \cup O^{\mathrm{c}}}, \tag{16}$$

where $R(E)$ is the regulation functional and α is the regulation parameter. The regulation penalty $R(E)$, which ensures that the extrapolation problem (16) is well posed, is usually determined by the *a priori* information about model error $E(\boldsymbol{x},t)$.

For obtaining higher efficiency, we here only need to extrapolate $E(\boldsymbol{x},t)$ in the forecast period without changing its value in the historical period. If we insert Eq. (15) into the objective function of expression (16), we obtain the following optimization problem:

$$\min_{E|_O} R(E)|_{O \cup O^{\mathrm{c}}}, \tag{17}$$

where $E|_O$ means that we only need to find $E(\boldsymbol{x},t)$ in the forecast period. Problem (17) means that the extrapolation value of model error in the forecast period can be extracted by its historical information.

Considering the analogue phenomenon of the atmosphere, we assume that the whole error data set $E(\boldsymbol{x},t)$ in historical and forecast periods exhibits self-similarity. For more reasonable treatment, we should divide the historical period O^{c} and forecast period O into smaller patches according to the regional characteristics of $E(\boldsymbol{x},t)$ that we already know,[34,35] and then, using the self-similarity of the atmosphere, we can obtain the extrapolation value by comparing the recent information about model error patches in the forecast period with the value of the patches in the historical period (see Fig. 1). This means that, by using nonlocal regularization theory,[36] we should construct the penalty term $R(E)|_{O \cup O^{\mathrm{c}}}$ in problem (17) as the following weighted sum of patch pairs of dataset $E(\boldsymbol{x},t)$:

$$R(E)|_{O \cup O^{\mathrm{c}}} = \sum_{\boldsymbol{X} \in O} \sum_{\boldsymbol{X}' \in O^{\mathrm{c}}} w(\boldsymbol{X}, \boldsymbol{X}') \left\| P_E(\boldsymbol{X}) - P_{\hat{E}}(\boldsymbol{X}') \right\|^2, \tag{18}$$

where

$$P_E(\boldsymbol{X}) = E(\boldsymbol{x} + p_{\boldsymbol{X}}, t_K + p_t); \;\; \boldsymbol{x} \in D, p_{\boldsymbol{X}} \in B_{\boldsymbol{x}}, p_t \in B_t, \tag{19}$$

and

$$\begin{aligned} &P_{\hat{E}}(\boldsymbol{X}') = E(\boldsymbol{x}' + p_{\boldsymbol{X}}, t' + p_t); \\ &\boldsymbol{x}' \in D, t' \in (t_0, t_K), p_{\boldsymbol{X}} \in B_{\boldsymbol{x}}, p_t \in B_t, \end{aligned} \tag{20}$$

denote the error patches of $E(\boldsymbol{x},t)$ in the forecast and historical periods, respectively. Here $B_{\boldsymbol{x}} = [-\delta\boldsymbol{x}, \delta\boldsymbol{x}]$, $B_t = [-\delta t, \delta t]$, and $\delta\boldsymbol{x}$, δt denote the space-scale and time-scale of $P_E(\boldsymbol{X})$

and $P_{\hat{E}}(\boldsymbol{X}')$, respectively. $w(\boldsymbol{X},\boldsymbol{X}')$ is a weight function to measure the similarity between $P_E(\boldsymbol{X})$ and $P_{\hat{E}}(\boldsymbol{X}')$. We can choose[36]

$$w(\boldsymbol{X},\boldsymbol{X}') = \exp(-\left\|P_E(\boldsymbol{X}) - P_{\hat{E}}(\boldsymbol{X}')\right\|^2/\sigma^2), \quad (21)$$

where we promote the averaging of similar patches and minimize the averaging of dissimilar patches. The parameter σ is a constant related to the convex distance metrics of patches $P_E(\boldsymbol{X})$ and $P_{\hat{E}}(\boldsymbol{X}')$.

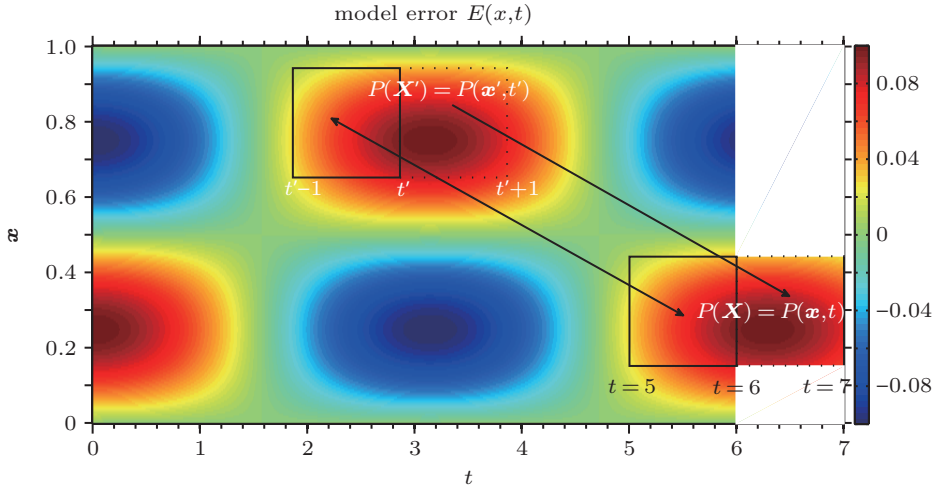


Fig. 1. (color online) Illustration of the proposed extrapolation procedure for model error in the forecast period: if the value of patch $P_E(\boldsymbol{X})$ in the recent time interval $5 \le t \le 6$ is similar to the value of patch $P_{\hat{E}}(\boldsymbol{X}')$ in the historical interval $t'-1 \le t \le t'$, then the extrapolation value of $P_E(\boldsymbol{X})$ in the forecast interval $6 \le t \le 7$ is also similar to the value of $P_{\hat{E}}(\boldsymbol{X}')$ in the historical interval $t' \le t \le t'+1$.

Equation (18) means that if the model error in the recent time interval $t_K - \delta t \le t \le t_K$ is similar to the value in the historical interval $t' - \delta t \le t \le t'$, the extrapolation value of model error in the forecast interval $t_K \le t \le t_K + \delta t$ is also similar to the value in the historical interval $t' \le t \le t' + \delta t$ by the continuous evolution of the atmosphere (see Fig. 1). The minimum solutions of problems (17) and (18) are the extrapolation values of model error in the forecast period.

3.2. Algorithm

Now we present a simple and effective iterative algorithm for problems (17) and (18). According to Eqs. (19), (20), and (21), the objective function of expression (18) can be written as

$$\begin{aligned} R(E)|_{O\cup O^c} &= \sum_{\boldsymbol{X}\in O}\sum_{\boldsymbol{X}'\in O^c} w(\boldsymbol{X},\boldsymbol{X}') \\ &\quad \times \sum_{p_{\boldsymbol{X}}\in B_{\boldsymbol{x}}, p_t\in B_t} |E(\boldsymbol{x}+p_{\boldsymbol{X}}, t_K+p_t) \\ &\quad - \hat{E}(\boldsymbol{x}'+p_{\boldsymbol{X}}, t+p_t)|^2 \\ &= \sum_{\boldsymbol{X}\in O}\sum_{\boldsymbol{X}'\in O^c} \gamma_E(\boldsymbol{X},\boldsymbol{X}')|E(\boldsymbol{x},t_K) - \hat{E}(\boldsymbol{x}',t)|^2. \end{aligned} \quad (22)$$

We use a change of variables $\boldsymbol{x} = \boldsymbol{x} + p_{\boldsymbol{X}}$, $t_K = t_K + p_t$, and $\boldsymbol{x}' = \boldsymbol{x}' + p_{\boldsymbol{X}}$, $t = t + p_t$ to derive the second step in Eq. (22).[36] The weight γ_E is specified by

$$\gamma_E(\boldsymbol{X},\boldsymbol{X}') = \sum_{\substack{p=(p_{\boldsymbol{X}},p_t)\\ p_{\boldsymbol{X}}\in B_{\boldsymbol{x}}, p_t\in B_t}} w(\boldsymbol{X}-p, \boldsymbol{X}'-p), \quad (23)$$

which denotes the sum of the similarity measures $w(\boldsymbol{X},\boldsymbol{X}')$ between all the patch pairs that contain $\boldsymbol{X}$ and $\boldsymbol{X}'$. Then we obtain an equivalent form of optimization problem (17) as follows:

$$\min_{\boldsymbol{E}} \boldsymbol{E}^{\mathrm{T}} A(\boldsymbol{E})\boldsymbol{E} - 2b^{\mathrm{T}}\boldsymbol{E}, \quad (24)$$

where $\boldsymbol{E}$ is the vector of $E(\boldsymbol{x},t)$ in the forecast period, $A(\boldsymbol{E})$ is the diagonal matrix with the entries $\gamma_E(\boldsymbol{X},\boldsymbol{X}')$ defined by Eq. (23), and b is the vector with the elements $\gamma_E(\boldsymbol{X},\boldsymbol{X}')\hat{E}(\boldsymbol{x}',t)$.

If we denote the objective function in expression (24) by $J(\boldsymbol{E}) = \boldsymbol{E}^{\mathrm{T}}A(\boldsymbol{E})\boldsymbol{E} - 2b^{\mathrm{T}}\boldsymbol{E}$, then we have

$$\min_{\boldsymbol{E}} J(\boldsymbol{E}). \quad (25)$$

The optimization problem (25) is actually a special case of problem (13) in Section 2, and we can compute it using Algorithm 1. For convenience, we also construct the following iterative method. The numerical experiments in this study demonstrate that the following iterative method is effective.

Algorithm 2 (iterative method for optimization problem (24)).

A-2-1. Given E_0 and $gtol > 0$. Set $k = 0$.

A-2-2. Compute $A(\boldsymbol{E}_k)$.

A-2-3. Solve equation $A(\boldsymbol{E}_k)\boldsymbol{E}_{k+1} = b$ for $\boldsymbol{E}_{k+1}$.

A-2-4. If $\|\boldsymbol{E}_{k+1} - \boldsymbol{E}_k\| / \|\boldsymbol{E}_k\| < gtol$ stop.

A-2-5. Let $k = k+1$. Goto A-2-2.

4. Numerical experiment

In Sections 2 and 3, we theoretically estimate and extrapolate model error in historical and forecast periods by constructing two inverse problem models. In this section, we will numerically demonstrate the effectiveness of our inverse

problem algorithms by using Burgers' equation. The Burgers' equation can be viewed as the Navier–Stokes equations in one space dimension[33], and it is usually written as

$$\begin{cases} \dfrac{\partial u}{\partial t} + u\dfrac{\partial u}{\partial x} - \nu\dfrac{\partial^2 u}{\partial x^2} = E(x,t), \quad 0 < x < 1, 0 < t < T, \\ u|_{t=0} = u_0(x), \end{cases} \tag{26}$$

where ν is the viscosity parameter and $u_0(x)$ is the initial condition. $u(x,t)$ and $E(x,t)$ are the state variable and forcing term, respectively.

Since we only consider the effect of model error on prediction results, for convenience, we take the zero boundary condition and accurate initial condition as $u_0(x) = \sin 2\pi x, x \in [0,1]$ in the following discussion. We suppose that equation (26) is the accurate model if the forcing term is

$$E(x,t) = A\cos t \sin 2\pi x, \quad x \in [0,1], 0 < t < T, \tag{27}$$

and its corresponding numerical model is as follows:

$$\begin{cases} \dfrac{\partial u}{\partial t} + u\dfrac{\partial u}{\partial x} - \nu\dfrac{\partial^2 u}{\partial x^2} = 0, \quad 0 < x < 1, \quad 0 < t < T, \\ u|_{t=0} = u_0(x), \end{cases} \tag{28}$$

where the model error term $E(x,t)$ is the actual forcing term in the accurate model (26), which cannot be described by the numerical model (28).

To illustrate the effectiveness of computing forcing term $E(x,t)$ in Eq. (26) by using our inverse problem algorithms, we first solve the accurate problem (26) with expression (27) to produce the 'observations' $u|_{t=t_1} = u_1, \ldots, u|_{t=t_K} = u_K$, and then we obtain the historical error term $E(x,t)$, according to Section 2, by solving the following optimization problem:

$$\min_E \frac{1}{2}\int_0^1 \sum_{k=1}^{K} |u(x,t_k) - u_k(x)|^2 \,\mathrm{d}x + \frac{\alpha}{2}\int_0^T \int_0^1 |E(x,t)|^2 \,\mathrm{d}x\mathrm{d}t, \tag{29}$$

such that $u(x,t_k)$ $(k = 1, \ldots, K)$ and $E(x,t)$ satisfy model (26). Now we further divide the spatial interval $[0,1]$ into n subintervals $[x_{i-1}, x_i]$ $(i = 1, \ldots, n)$ with $x_i = ih$ and $h = 1/n$. We define the following piecewise linear functions as the basis functions in Eqs. (6) and (7):

$$\varphi_i(x) = \begin{cases} (x - x_{i-1})/h, & x \in [x_{i-1}, x_i] \cap [0,1], \\ (x_{i+1} - x)/h, & x \in [x_i, x_{i+1}] \cap [0,1], \\ 0, & \text{else}, \end{cases} \quad i = 0, \ldots, n,$$

which satisfies $\varphi_i(x_i) = 1$ and $\varphi_i(x_j) = 0, i \neq j$.

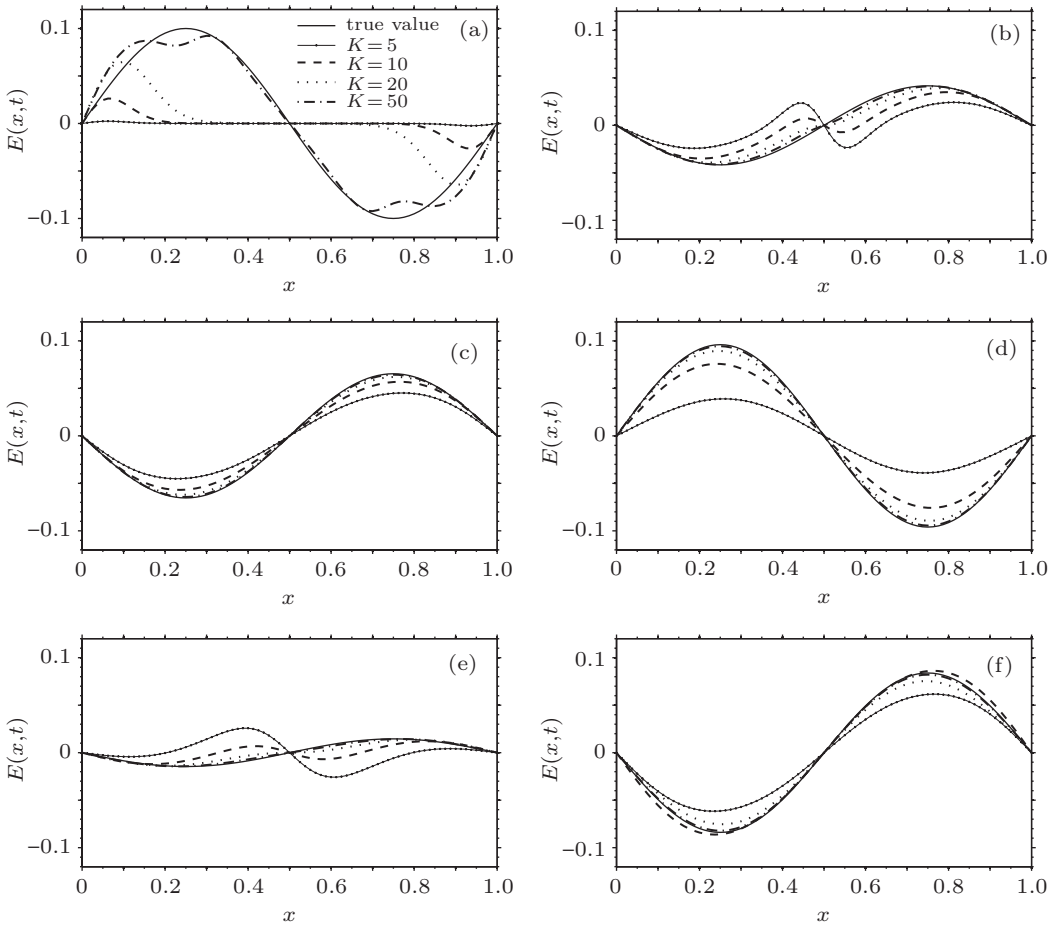


Fig. 2. Accurate values and numerical values of model error $\hat{E}(x,t)$ in the historical period ($0 \le t \le 10$, at times $t = 0$ (a), 2 (b), 4 (c), 6 (d), 8 (e), and 10 (f)). The solid line denotes the accurate value computed by Eq. (27), and dashed lines denote the numerical value. K denotes the number of the 'observations' used in the constrained optimization problems (26) and (29).

After we have obtained the historical dataset $E(x,t)$, we first need to compare $E(x,t)$ with its accurate form Eq. (27) to ensure the accuracy of our computing, and then we extrapolate it in the forecast period. We also need to compare the extrapolation $E(x,t)$ with its accurate form Eq. (27) in the forecast period to testify the effectiveness of extrapolation algorithms.

For convenience, we call the solutions of numerical model (28) 'forecasts' and the solutions of model (26) with our computed term $E(x,t)$ in both the historical period and the forecast period 'improved forecasts'. We take $T = 10$, time step $\delta t = 0.025$, space step $\delta x = 0.0125$, and parameter $\nu = 0.02$ in both Eq. (26) and Eq. (28). We choose parameter $\alpha = 0.05$ in expression (29), $\sigma^2 = 0.1$ in Eq. (21), and $A = 0.1$ in Eq. (27). We compute the mean square error of state variable $u(x,t)$ according to

$$e(t_j) = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(u(x_i,t_j) - \tilde{u}(x_i,t_j))^2}, \tag{30}$$

where $u(x_i,t_j)$ denotes the 'forecast' or 'improved forecast' and $\tilde{u}(x_i,t_j)$ denotes the 'observation'.

Figure 2 shows comparisons between the true forcing term $E(x,t)$ (see Eq. (27)) and our computed value in the historical period ($0 \le t \le 10$). We notice that the computed $E(x,t)$ is much closer to its accurate value when more "observations" (see $K = 50$ in Fig. 2) are used in the constrained optimization problems (26) and (29). Figure 3 presents comparisons between the true forcing term $E(x,t)$ (also see Eq. (27)) and our extrapolation value in the forecast period ($10 \le t \le 11$). It is obvious that the extrapolation value can closely approximate the true value when the historical model error is well computed (see $K = 50$ in Fig. 3). Figures 2 and 3 illustrate that our inverse problem procedures proposed in Sections 2 and 3 can well recover the unknown model error term $E(x,t)$ both in the historical period and in the forecast period.

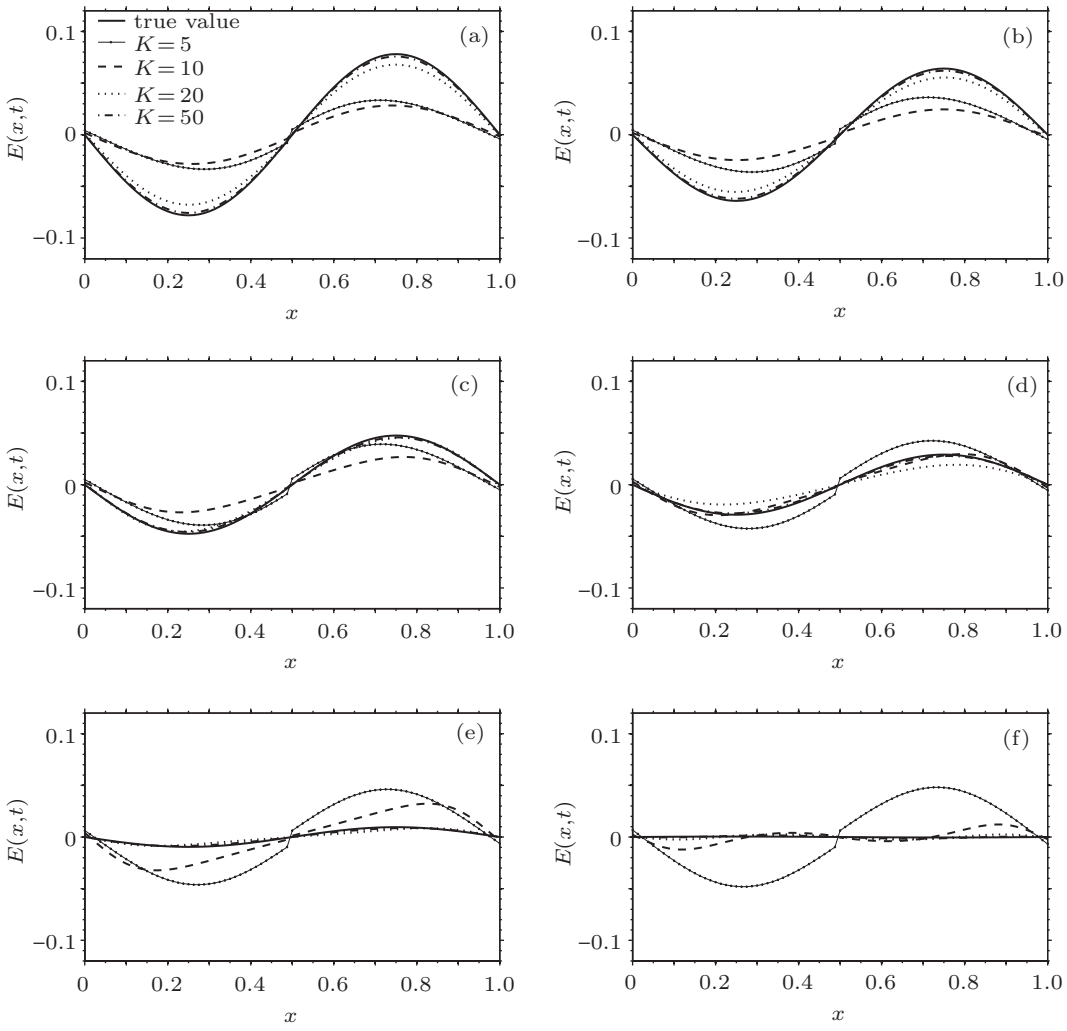


Fig. 3. Accurate values and extrapolation values of model error $\hat{E}(x,t)$ in the forecast period ($10 \le t \le 11$, at times $t = 10.1$ (a), 10.3 (b), 10.5 (c), 10.7 (d), 10.9 (e), and 11 (f)). The solid line denotes the accurate value computed from Eq. (27), and dashed lines denote the numerical value. K denotes the number of 'observations' used in the constrained optimization problems (26) and (29). Here, we divide the forecast period O into two smaller patches in the model error extrapolation problems (17) and (18).

After finding the model error term $E(x,t)$ in the historical period and the forecast period, we insert it into model (26) and obtain the 'improved forecasts' for the state variable $u(x,t)$. Figure 4 presents the mean square errors of 'forecasts' and 'improved forecasts' in the historical period ($0 \le t \le 10$) and forecast period ($10 \le t \le 11$). It shows that the improved model (26) has higher forecast skill than the numerical model (28). The improved forecast error is far less than the forecast error if K and the patch in the forecast period O become bigger (see Fig. 4(a)), which implies that the more recent observations are used in our inverse problem methods, the better the improvement of forecasts is.

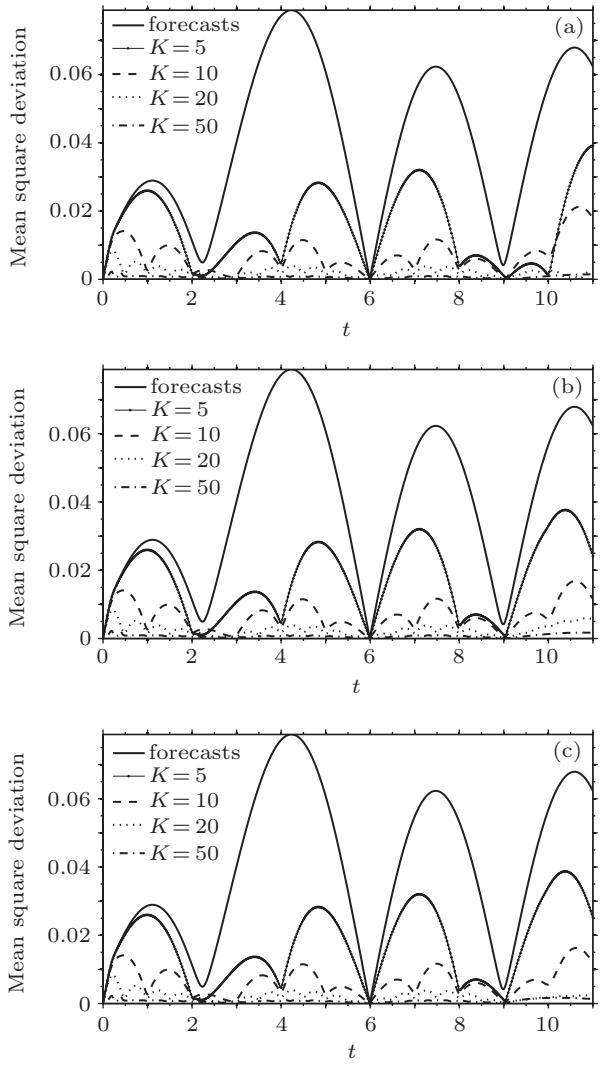


Fig. 4. Mean square errors of 'forecasts' (solid line) and their 'improved forecasts' (dashed lines) in both the historical period ($0 \le t \le 10$) and the forecast period ($10 \le t \le 11$): we divide the forecast period O into (a) one patch, (b) two smaller patches, and (c) four smaller patches in the model error extrapolation problems (17) and (18).

5. Conclusions

In this study, we show a novel method of estimating and extrapolating model error using inverse problem ideas. We propose to obtain the model error in both the historical period and forecast period by solving two inverse problems using observations and self-similarity of the atmosphere. The numerical experiments demonstrate the substantial forecast improvement resulting from our inverse problem algorithms. In particular, the inverse problem ideas in this paper can be used in a global model, and especially in a local NWP model. For example, researchers are now increasingly concerned about the accuracy of timing, quantitative and fixed location prediction. If we use the local observations of the investigated phenomenon in the inverse problem methods to correct the local numerical model, this will surely improve the prediction for the investigated phenomenon. This will lead to using large scale past data more effectively and economically.

Appendix A

Following Refs. [33] and [37], we present the detailed procedures to compute the gradients $\nabla\hat{f}(\boldsymbol{E}_k)$ and Hessians $\nabla^2\hat{f}(\boldsymbol{E}_k)$ in Algorithm 1. Because $\psi = \psi(\boldsymbol{E})$ is determined by the nonlinear equation $c(\psi,\boldsymbol{E}) = 0$, the Jacobian of $\psi = \psi(\boldsymbol{E})$ is satisfied as shown below:

$$c_\psi(\psi,\boldsymbol{E})|_{\psi=\psi(\boldsymbol{E})}\psi_{\boldsymbol{E}}(\boldsymbol{E}) + c_{\boldsymbol{E}}(\psi,\boldsymbol{E})|_{\psi=\psi(\boldsymbol{E})} = 0.$$

To simplify the notation, we write $c_\psi(\psi(\boldsymbol{E}),\boldsymbol{E})$ and $c_{\boldsymbol{E}}(\psi(\boldsymbol{E}),\boldsymbol{E})$ instead of $c_\psi(\psi,\boldsymbol{E})|_{\psi=\psi(\boldsymbol{E})}$ and

$$c_{\boldsymbol{E}}(\psi,\boldsymbol{E})|_{\psi=\psi(\boldsymbol{E})},$$

respectively. We have

$$\psi_{\boldsymbol{E}}(\boldsymbol{E}) = -c_\psi(\psi(\boldsymbol{E}),\boldsymbol{E})^{-1}c_{\boldsymbol{E}}(\psi(\boldsymbol{E}),\boldsymbol{E}). \tag{A1}$$

Since $\hat{f}(\boldsymbol{E}) = f(\psi(\boldsymbol{E}),\boldsymbol{E})$, the gradient $\nabla\hat{f}(\boldsymbol{E})$ is given by

$$\nabla\hat{f}(\boldsymbol{E}) = \psi_{\boldsymbol{E}}(\boldsymbol{E})^{\mathrm{T}}\nabla_\psi f(\psi(\boldsymbol{E}),\boldsymbol{E}) + \nabla_{\boldsymbol{E}} f(\psi(\boldsymbol{E}),\boldsymbol{E}). \tag{A2}$$

If we insert Eq. (A1) into Eq. (A2), then we obtain

$$\nabla\hat{f}(\boldsymbol{E}) = -c_{\boldsymbol{E}}(\psi(\boldsymbol{E}),\boldsymbol{E})^{\mathrm{T}}c_\psi(\psi(\boldsymbol{E}),\boldsymbol{E})^{-\mathrm{T}}\nabla_\psi f(\psi(\boldsymbol{E}),\boldsymbol{E}) + \nabla_{\boldsymbol{E}} f(\psi(\boldsymbol{E}),\boldsymbol{E}). \tag{A3}$$

In order to compute the gradient more efficiently, we define

$$\lambda(\boldsymbol{E}) = -c_\psi(\psi(\boldsymbol{E}),\boldsymbol{E})^{-T}\nabla_\psi f(\psi(\boldsymbol{E}),\boldsymbol{E}),$$

or, equivalently, we can define $\lambda(\boldsymbol{E})$ as the solution of

$$c_\psi(\psi(\boldsymbol{E}),\boldsymbol{E})^{\mathrm{T}}\lambda(\boldsymbol{E}) = -\nabla_\psi f(\psi(\boldsymbol{E}),\boldsymbol{E}). \tag{A4}$$

With this quantity, the gradient can now be written as

$$\nabla\hat{f}(\boldsymbol{E}) = \nabla_{\boldsymbol{E}} f(\psi(\boldsymbol{E}),\boldsymbol{E}) + c_{\boldsymbol{E}}(\psi(\boldsymbol{E}),\boldsymbol{E})^{\mathrm{T}}\lambda(\boldsymbol{E}). \tag{A5}$$

Algorithm A1 (gradient computation).

A1-1. Given $\boldsymbol{E}$, solve $c(\psi,\boldsymbol{E}) = 0$ for $\psi(\boldsymbol{E})$.

A1-2. Solve Eq. (A4) for $\lambda(\boldsymbol{E})$.

A1-3. Compute $\nabla\hat{f}(\boldsymbol{E}) = \nabla_{\boldsymbol{E}} f(\psi(\boldsymbol{E}), \boldsymbol{E}) + c_{\boldsymbol{E}}(\psi(\boldsymbol{E}), \boldsymbol{E})^{\mathrm{T}}\lambda(\boldsymbol{E})$.

Now we need to compute the Hessian $\nabla^2\hat{f}(\boldsymbol{E})$ for given $\boldsymbol{E}$ from Eq. (A5). In fact, the Lagrangian, corresponding to the constraint problem (12), is as follows:

$$L(\psi, \boldsymbol{E}, \lambda) = f(\psi, \boldsymbol{E}) + \lambda^{\mathrm{T}} c(\psi, \boldsymbol{E}); \tag{A6}$$

then equation (A4) can be written as

$$\nabla_{\psi} L(\psi(\boldsymbol{E}), \boldsymbol{E}, \lambda(\boldsymbol{E})) = 0. \tag{A7}$$

Differentiating Eq. (A7) gives

$$\begin{aligned}&\nabla_{\psi\psi} L(\psi(\boldsymbol{E}), \boldsymbol{E}, \lambda(\boldsymbol{E}))\psi_{\boldsymbol{E}}(\boldsymbol{E}) + \nabla_{\psi\boldsymbol{E}} L(\psi(\boldsymbol{E}), \boldsymbol{E}, \lambda(\boldsymbol{E}))\\ &+ \nabla_{\psi\lambda} L(\psi(\boldsymbol{E}), \boldsymbol{E}, \lambda(\boldsymbol{E}))\lambda_{\boldsymbol{E}}(\boldsymbol{E}) = 0.\end{aligned}$$

If we use $\nabla_{\psi\lambda} L(\psi, \boldsymbol{E}, \lambda) = c_{\psi}(\psi, \boldsymbol{E})^{\mathrm{T}}$ and Eq. (A1) in the above equation, we have

$$\begin{aligned}\lambda_{\boldsymbol{E}}(\boldsymbol{E}) = c_{\psi}(\psi, \boldsymbol{E})^{-T} [&\nabla_{\psi\psi} L(\psi(\boldsymbol{E}), \boldsymbol{E}, \lambda(\boldsymbol{E}))\\ &\times c_{\psi}(\psi(\boldsymbol{E}), \boldsymbol{E})^{-1} c_{\boldsymbol{E}}(\psi(\boldsymbol{E}), \boldsymbol{E})\\ &- \nabla_{\psi\boldsymbol{E}} L(\psi(\boldsymbol{E}), \boldsymbol{E}, \lambda(\boldsymbol{E}))].\end{aligned} \tag{A8}$$

Moreover, equation (A5) can be written as

$$\nabla\hat{f}(\boldsymbol{E}) = \nabla_{\boldsymbol{E}} L(\psi, \boldsymbol{E}, \lambda)|_{\psi=\psi(\boldsymbol{E}), \lambda=\lambda(\boldsymbol{E})}. \tag{A9}$$

Now we can compute the Hessian $\nabla^2\hat{f}(\boldsymbol{E})$ by differentiating Eq. (A9), and have

$$\begin{aligned}\nabla^2\hat{f}(\boldsymbol{E}) = &\nabla_{\boldsymbol{E}\psi} L(\psi(\boldsymbol{E}), \boldsymbol{E}, \lambda(\boldsymbol{E}))\psi_{\boldsymbol{E}}(\boldsymbol{E})\\ &+ \nabla_{\boldsymbol{E}\boldsymbol{E}} L(\psi(\boldsymbol{E}), \boldsymbol{E}, \lambda(\boldsymbol{E}))\\ &+ \nabla_{\boldsymbol{E}\lambda} L(\psi(\boldsymbol{E}), \boldsymbol{E}, \lambda(\boldsymbol{E}))\lambda_{\boldsymbol{E}}(\boldsymbol{E}).\end{aligned} \tag{A10}$$

If we insert Eqs. (A1) and (A8) into Eq. (A10) and use $\nabla\boldsymbol{E}\lambda L(\psi, \boldsymbol{E}, \lambda) = c_{\boldsymbol{E}}(\psi, \boldsymbol{E})^{\mathrm{T}}$, the Hessian $\nabla^2\hat{f}(\boldsymbol{E})$ can be written as

$$\begin{aligned}&\nabla^2\hat{f}(\boldsymbol{E})\\ &= c_{\boldsymbol{E}}(\psi, \boldsymbol{E})^{\mathrm{T}} c_{\psi}(\psi, \boldsymbol{E})^{-\mathrm{T}} \nabla_{\psi\psi} L(\psi(\boldsymbol{E}), \boldsymbol{E},\\ &\lambda(\boldsymbol{E})) c_{\psi}(\psi(\boldsymbol{E}), \boldsymbol{E})^{-1} c_{\boldsymbol{E}}(\psi(\boldsymbol{E}), \boldsymbol{E})\\ &- c_{\boldsymbol{E}}(\psi, \boldsymbol{E})^{\mathrm{T}} c_{\psi}(\psi, \boldsymbol{E})^{-\mathrm{T}} \nabla_{\psi\boldsymbol{E}} L(\psi(\boldsymbol{E}), \boldsymbol{E}, \lambda(\boldsymbol{E}))\\ &- \nabla_{\boldsymbol{E}\psi} L(\psi(\boldsymbol{E}), \boldsymbol{E}, \lambda(\boldsymbol{E})) c_{\psi}(\psi(\boldsymbol{E}), \boldsymbol{E})^{-1} c_{\boldsymbol{E}}(\psi(\boldsymbol{E}), \boldsymbol{E})\\ &+ \nabla_{\boldsymbol{E}\boldsymbol{E}} L(\psi(\boldsymbol{E}), \boldsymbol{E}, \lambda(\boldsymbol{E})).\end{aligned} \tag{A11}$$

Obviously equation (A11) can be used to compute the Hessian $\nabla^2\hat{f}(\boldsymbol{E})$. However, in the Newton-CG algorithm 1, we only require the Hessian-times-vector products $\nabla^2\hat{f}(\boldsymbol{E})v$. Using Eq. (A11) we have the following algorithm.

Algorithm A2 (Hessian-times-vector computation).

A2-1. Given $\boldsymbol{E}$, solve $c(\psi, \boldsymbol{E}) = 0$ for $\psi(\boldsymbol{E})$.

A2-2. Solve Eq. (A4) for $\lambda(E)$.

A2-3. Solve equation $c_{\psi}(\psi(\boldsymbol{E}), \boldsymbol{E})w = c_{\boldsymbol{E}}(\psi(\boldsymbol{E}), \boldsymbol{E})v$ for w.

A2-4. Solve equation

$$\begin{aligned}c_{\psi}(\psi, \boldsymbol{E})^{\mathrm{T}} p = &\nabla_{\psi\psi} L(\psi(\boldsymbol{E}), \boldsymbol{E}, \lambda(\boldsymbol{E}))w\\ &- \nabla_{\psi\boldsymbol{E}} L(\psi(\boldsymbol{E}), \boldsymbol{E}, \lambda(\boldsymbol{E}))v \quad \text{for} \quad p.\end{aligned}$$

A2-5. Compute

$$\begin{aligned}\nabla^2\hat{f}(\boldsymbol{E})v = &c_{\boldsymbol{E}}(\psi, \boldsymbol{E})^{\mathrm{T}} p - \nabla_{\boldsymbol{E}\psi} L(\psi(\boldsymbol{E}), \boldsymbol{E}, \lambda(\boldsymbol{E}))w\\ &+ \nabla_{\boldsymbol{E}\boldsymbol{E}} L(\psi(\boldsymbol{E}), \boldsymbol{E}, \lambda(\boldsymbol{E}))v.\end{aligned}$$

References

[1] Toth Z and Kalnay E 1993 *Bull. Am. Meteorol. Soc.* **74** 2317
[2] Molteni F, Buizza R, Palmer T N and Petroliagis T 1996 *Q. J. R. Meteorol. Soc.* **122** 73
[3] Anderson J L 2001 *Mon. Wea. Rev.* **129** 2884
[4] Whitaker J S and Hamill T M 2002 *Mon. Wea. Rev.* **130** 1913
[5] Danforth C M and Yorke J A 2006 *Phys. Rev. Lett.* **96** 144102
[6] Zheng Z H, Feng G L, Huang J P and Chou J F 2012 *Acta Phys. Sin.* **61** 199203 (in Chinese)
[7] Feng G L and Dong W J 2003 *Acta Phys. Sin.* **52** 2347 (in Chinese)
[8] Danforth C M and Kalnay E 2008 *Geophys. Res. Lett.* **35** L24805
[9] Wang P and Dai X G 2005 *Acta Phys. Sin.* **54** 4961 (in Chinese)
[10] Danforth C M, Kalnay E and Takemasa M 2007 *Mon. Wea. Rev.* **135** 281
[11] Leith C E 1978 *Ann. Rev. Fluid Mech.* **10** 107
[12] Ferranti L, Klinker E, Hollingsworth A and Hoskins B J 2002 *Quart. J. Roy. Meteor. Soc.* **128** 1623
[13] Shao A M, Xi S and Qiu C J 2009 *Sci. China: Ser. D* **52** 1650
[14] Kaas E, Guldberg A, May W and Deque M 1999 *Tellus* **51** 612
[15] Gu Z C 1958 *Acta Meteor. Sin.* **29** 176 (in Chinese)
[16] Chou J F 1974 *Sci. Sin.* **17** 814 (in Chinese)
[17] Qiu C J and Chou J F 1988 *Sci. Atmos. Sin.* **12** 225 (in Chinese)
[18] Qiu C J and Chou J F 1989 *Sci. Atmos. Sin.* **13** 22 (in Chinese)
[19] Qiu C J and Chou J F 1990 *Sci. China: Ser. B* **20** 218
[20] Huang J P and Yi Y 1991 *Sci. China: Ser. B* **34** 1246
[21] Huang J P, Yi Y H, Wang S W and Chou J F 1993 *Q. J. R. Meteorol. Soc.* **119** 547
[22] Bao M, Ni Y Q and Chou J F 2004 *Chin. Sci. Bull.* **49** 1296
[23] Ren H L, Zhang P Q, Li W J and Chou J F 2006 *Acta Phys. Sin.* **55** 4388 (in Chinese)
[24] Zheng Z H, Huang J P, Feng G L and Chou J F 2013 *Sci. China: Ser. D* **56** 878
[25] Zupanski D and Zupanski M 2006 *Mon. Wea. Rev.* **134** 1337
[26] Gillijns S and Moor B D 2007 *Nonlinear Process. Geophys.* **14** 59
[27] Carrassi A and Vannitsem S 2010 *Mon. Wea. Rev.* **138** 3369
[28] Feng F and Duan W S 2013 *Sci. China: Earth Sci.* **56** 434
[29] Xue H L, Shen X S and Chou J F 2013 *Adv. Atmos. Sci.* **30** 1249
[30] Duan W S and Zhou F F 2013 *Tellus* **65** 18452
[31] Duan W S, Tian B and Xu H 2013 *Clim. Dyn.* doi: 10.1007s00382-013-1993-4
[32] Doicu A, Trautmann T and Schreier F 2010 *Numerical Regularization for Atmospheric Inverse Problem* (Berlin: Springer) pp. 163–217
[33] Heinkenschloss M 2012 *Technical Report TR*08-05, Rice University
[34] Zhang Z S, Gong Z Q, Zhi R, Feng G L and Hu J G 2011 *Chin. Phys. B* **20** 019201
[35] Feng G L, Gong Z Q, Zhi R and Zhang D Q 2008 *Chin. Phys. B* **17** 2745
[36] Yang Z L and Jacob M 2013 *IEEE Trans. Image Process* **22** 3192
[37] Fikriye Yılmaz and Bülent Karasözen 2014 *J. Comput. Appl. Math.* **259** 771