Tellus (2005), 57A, 736–741
Printed in Singapore. 


TELLUS

# Initial condition and parameter estimation in physical 'on–off' processes by variational data assimilation

*By* JIAFENG WANG*†, MU MU and QIN ZHENG, *State Key Laboratory of Numerical Modeling for Atmospheric Sciences and Geophysical Fluid Dynamics (LASG), Institute of Atmospheric Physics, Chinese Academy of Sciences, Beijing, China*

(Manuscript received 29 November 2004; in final form 7 February 2005)

ABSTRACT

In this paper we investigate the feasibility of using the variational data assimilation method to estimate both the initial condition and parameter in an idealized simple model with parametrized discontinuity. A method based on the non-linear perturbation equation (NPE) is applied to calculate the gradient of the cost function with respect to the control variables used in the minimization procedure. The results obtained by this method and by the conventional treatment (ignoring the variation of the switch point due to the perturbation in initial condition, i.e. keeping the switching point in the tangent linear model the same as in the basic state) are compared through numerical experiment. Because the cost function could have artificial minima after time discretization, the optimization algorithm combined with the NPE method is less sensitive to the first guess in retrieving the optimal initial condition and parameter value. It is shown that the gradient computation deserves consideration when there are discontinuities in the model used in variational data assimilation.

## 1. Introduction

Precipitation data assimilation has received much attention in the demand of operational forecast (Fillion and Mahfouf, 2000; Marecal and Mahfouf, 2000). In precipitation processes, moist convection schemes are strongly non-linear and are characterized by the existence of 'on–off' switches. The application of the adjoint method in this case has been studied intensely (for example, Zou et al., 1993; Xu, 1996, 1998; Zou, 1997, 1998). Conventional treatment of the switch, which ignores the variation of the switch point due to the perturbation in initial condition (i.e. keeping the switching point in the tangent linear model the same as in the basic state, which is called the conventional method in the following), is feasible if the 'on–off' switch is time continuous. However, if the switch is discontinuous, it will bring in error in the calculated gradient of the cost function with respect to the initial condition. In Xu (1996), a generalized adjoint is proposed and the correct gradient can be obtained when the 'on–off' switch is not continuous. Further, a coarse-grain gradient is raised and applied to the Lorenz model (Xu et al., 1998; Xu and Gao, 1999).

In Mu and Wang (2003), the non-linear perturbation equation (NPE) method is introduced to deal with the discontinuous switch in variational assimilation. Numerical experiment shows that the gradient obtained from the NPE is helpful in finding the global minimum during minimization, even if there are artificial minima on the discrete cost function resulting from discretization. In fact, this gradient is exactly the discrete form of its analytical expression. This approach is enlightened by the fact that the global minimum of the discrete cost function is close to the true minimum of the analytical cost function. Its philosophy is the same as the coarse-grain gradient in Xu et al. (1998) – the gradient used in minimization is not the exact gradient of the discrete cost function. However, we emphasize that it is realized in a different way.

Here, the NPE method is examined in estimating both the initial condition and parameter value of a simple model with a discontinuous switch.

The parameter estimation is closely related to the physical process scheme. In parametrization schemes, many physical characteristics of the state variables or the interaction between two processes are described by parameters. Traditionally, the parameters are determined by trial and error. Because they are elements of the model's formulation, there is no doubt that an improvement of the parameters is useful to enhance the numerical model's forecast skill.

With the ability to minimize the difference between the model state and observation data within the dynamical constraint,

*Corresponding author:
e-mail: jiafeng.wang@ec.gc.ca
†On partial leave of absence to the Climate Monitoring and Data Interpretation Division (CCRM), Climate Research Branch (CRB), Meteorological Service of Canada, 4905 Dufferin Street, Downsview, Ontario, Canada, M3H 5T4.

variational data assimilation can also be used in parameter estimation. A detailed survey of the state of the art of parameter estimation in meteorology and oceanography was given by Navon (1997), in which total variation regularization techniques for parameter estimation with discontinuities were also reviewed. Lu and Hsieh (1998) compared the differences between both retrieving the parameters and initial conditions together and retrieving either of them separately. Zhu and Navon (1999) estimated the physical and numerical parameters as well as the initial condition in the Florida State University (FSU) spectral model and displayed their effects on the ensuing forecast. Although the above work focused on complex models, the topic of the impact of discontinuous 'on–off' switches on the adjoint optimal determination of the parameter value along with the initial condition has not yet been mentioned. Also, it is also worthwhile examining the performance of the NPE method on such a problem. This is the aim of our present study.

The structure of this paper is as follows. In the next section, the NPE method is applied to an extended version of the idealized simple 'on–off' switch model originally developed by Xu (1996) and adopted by Zou (1997) with modifications. The procedure for both the initial condition and parameter estimation in a certain conceptual model is presented in Section 3. The computational details and numerically retrieved results from the optimization algorithm with the conventional and NPE methods are shown in Section 4. Section 5 contains a summary and discussion.

## 2. Gradient calculation based on the non-linear perturbation equation method

We consider a general model including an 'on–off' switch:

$$\begin{aligned} \frac{\mathrm{d}q}{\mathrm{d}t} &= F(t) + G(t, q, \alpha)H_+(q - q_c) \qquad 0 \le t \le T \\ q|_{t=0} &= q_0. \end{aligned} \tag{2.1}$$

This model is an extended version of the idealized model of Xu (1996) adopted by several others (Zou, 1997; Mu and Wang, 2003). Here, $q$ is the model variable, $q_0$ is the initial value, $q_c$ is the threshold, and $\alpha$ is the parameter. $H(\cdot)$ is the Heaviside unit-step function. We assume that $F(t) \ge \delta > 0$ and that the first-order differential of $F(t)$ and all second-order partial differentials of $G(t, q, \alpha)$ are continuous. To ensure that the switch turns on once in the time window [0, $T$], we assume (i) $q_0 < q_c < q_0 + \delta T$ and (ii) $F(t) > -G(t, q, \alpha)$.

It can be proved that eq. (2.1) is well posed in [0, $T$]. To estimate the two control variables ($q_0$ and $\alpha$) simultaneously, the cost function is taken as

$$J(q_0, \alpha) = \frac{1}{2}\int_0^T [q(t) - q_{\mathrm{obs}}(t)]^2 \,\mathrm{d}t, \tag{2.2}$$

where $q(t)$ is the solution to eq. (2.1), which is continuous in [0, $T$], and $q_{\mathrm{obs}}(t)$ is the observation of $q(t)$.

It can be proved that there exists unique $\tau$, $0 < \tau < T$, satisfying $q(\tau) = q_c$. Furthermore, $q(t) < q_c$ if $0 \le t < \tau$, and $q(t) \ge q_c$ otherwise. The time $\tau$ is called the 'on–off' switch time of problem (2.1).

In the following, we use the NPE method to calculate the gradient of the cost function (2.2) with respect to the two control variables.

Let the perturbations of the initial value and parameter be $q_0'$ and $\alpha'$, respectively. When $q_0' > 0$, which satisfies $q_0 + q_0' < q_c$, the switch time changes from $\tau$ to $\tau'$, and

$$\tau - \tau' = \frac{q_0'}{F(\tau)} + O\left(q_0'^2\right). \tag{2.3}$$

Here we have applied the continuity of $F'(t)$ in [0, $T$] and $F(t) \ge \delta > 0$.

The non-linear perturbation equation of (2.1) is

$$\begin{aligned} \frac{\mathrm{d}q'}{\mathrm{d}t} &= 0 \qquad 0 \le t < \tau' \\ \frac{\mathrm{d}q'}{\mathrm{d}t} &= G(t, q + q', \alpha + \alpha') \qquad \tau' \le t < \tau \\ \frac{\mathrm{d}q'}{\mathrm{d}t} &= G(t, q + q', \alpha + \alpha') - G(t, q, \alpha) \qquad \tau \le t \le T \\ q'|_{t=0} &= q_0'. \end{aligned} \tag{2.4}$$

It can be proved that the solution to perturbation eq. (2.4) satisfies

$$q'(t) = O(q_0', \alpha'). \tag{2.5}$$

As in Le Dimet and Talagrand (1986), introducing a new variable $q^*$, multiplying eq. (2.4) by $q^*$ and integrating it in [0, $T$], we obtain

$$\begin{aligned} \int_0^T \frac{\mathrm{d}q^*}{\mathrm{d}t} q' \,\mathrm{d}t &= q^*(T)q'(T) - q^*(0)q'(0) \\ &\quad - \int_{\tau'}^{\tau} G(t, q + q', \alpha + \alpha')q^* \,\mathrm{d}t - \int_{\tau}^{T} \\ &\quad \times [G(t, q + q', \alpha + \alpha') - G(t, q, \alpha)]q^* \,\mathrm{d}t. \end{aligned} \tag{2.6}$$

It can be proved that

$$\int_{\tau'}^{\tau} G(t, q + q', \alpha + \alpha')q^* \,\mathrm{d}t = G(\tau, q_c, \alpha)q^*(\tau)\frac{q_0'}{F(\tau)} + \mathrm{O}\left(q_0'^2, q_0'\alpha'\right),$$

and

$$\int_{\tau}^{T} [G(t, q + q', \alpha + \alpha') - G(t, q, \alpha)]q^* \,\mathrm{d}t = \int_{\tau}^{T} \times \left[\frac{\partial G}{\partial q}(t, q, \alpha)q' + \frac{\partial G}{\partial \alpha}(t, q, \alpha)\alpha'\right] q^* \,\mathrm{d}t + \mathrm{O}\left(q_0'^2, q_0'\alpha', \alpha'^2\right).$$

Hence

$$\int_0^T \frac{\mathrm{d}q^*}{\mathrm{d}t} q'\,\mathrm{d}t = q^*(T)q'(T) - q^*(0)q'(0) - \frac{G(\tau, q_c, \alpha)}{F(\tau)} q^*(\tau)q_0'$$
$$- \int_\tau^T \left[\frac{\partial G}{\partial q}(t, q, \alpha)q' + \frac{\partial G}{\partial \alpha}(t, q, \alpha)\alpha'\right] q^*\,\mathrm{d}t$$
$$+ \mathrm{O}\left(q_0'^2, q_0'\alpha', \alpha'^2\right). \tag{2.7}$$

For the case of $q_0' < 0$, the same result can be obtained.

According to the definition of cost function (2.2), it can be derived that

$$\int_0^T (q - q^{\mathrm{obs}})q'\,\mathrm{d}t = \frac{\partial J}{\partial q_0} q_0' + \frac{\partial J}{\partial \alpha}\alpha' + o(q_0', \alpha').$$

If $q^*$ is the solution to the following equation

$$\frac{\mathrm{d}q^*}{\mathrm{d}t} = -(q - q^{\mathrm{obs}}) - \frac{\partial G}{\partial q}(t, q, \alpha)H_+(t - \tau)q^*, \quad 0 \le t \le T,$$
$$q^*|_{t=T} = 0, \tag{2.8}$$

then

$$\frac{\partial J}{\partial q_0} q_0' + \frac{\partial J}{\partial \alpha}\alpha' = q^*(0)q_0' + \frac{G(\tau, q_c, \alpha)}{F(\tau)} q^*(\tau)q_0'$$
$$+ \alpha' \int_\tau^T \frac{\partial G}{\partial \alpha}(t, q, \alpha)q^*\,\mathrm{d}t.$$

That is,

$$\frac{\partial J}{\partial q_0} = q^*(0) + \frac{G(\tau, q_c, \alpha)}{F(\tau)} q^*(\tau), \tag{2.9a}$$

$$\frac{\partial J}{\partial \alpha} = \int_\tau^T \frac{\partial G}{\partial \alpha}(t, q, \alpha)q^*\,\mathrm{d}t. \tag{2.9b}$$

It can be found that although eq. (2.8) is derived from the NPE approach, it is in the same form as the adjoint model derived through conventional treatment. Another point that should be mentioned is the additional term related to $q^*(\tau)$ in $\partial J/\partial q_0$ due to the existence of the 'on–off' switch (see eq. 2.9).

## 3. Initial condition and parameter estimation for a conceptual model

Consider a conceptual model (Zou, 1997)

$$\frac{\mathrm{d}q}{\mathrm{d}t} = F + H_+(q - q_c)\alpha q \qquad 0 \le t \le T$$
$$q|_{t=0} = q_0 \tag{3.1}$$

where the second source term is due to the parametrized process, and $F$ is the source term due to other processes. It is obvious that there is a discontinuous jump in the right-hand side of eq. (3.1).

The cost function $J$ is taken as eq. (2.2) and all the conditions required in eq. (2.1) are satisfied in eq. (3.1). From eq. (2.9), we obtain the formula for computing the gradients of $J$ with respect to $q_0$ and $\alpha$

$$\frac{\partial J}{\partial q_0} = q^*(0) + \frac{\alpha}{F} q_c q^*(\tau), \tag{3.2a}$$

$$\frac{\partial J}{\partial \alpha} = \int_\tau^T q q^*\,\mathrm{d}t, \tag{3.2b}$$

where $\tau$ is the switch time and $q^*$ is the solution to

$$-\frac{\mathrm{d}q^*}{\mathrm{d}t} = \alpha q^* + q - q_{\mathrm{obs}}, \ \tau \le t \le T$$
$$-\frac{\mathrm{d}q^*}{\mathrm{d}t} = q - q_{\mathrm{obs}}, \qquad 0 \le t < \tau \tag{3.3}$$

with $q^*(T) = 0$.

Using the conventional treatment on the switch, we obtain the conventional adjoint model, which is in the same form as eq. (3.3), but in the gradient expressions, there is a difference from eq. (3.2): no term related with $q^*(\tau)$ appears, i.e. $\partial J/\partial q_0 = q^*(0)$.

Once the observation data are given in time window [0, $T$], the general procedure for estimating the initial condition and parameter using the adjoint method is as follows. (1) Specify a first guess of $q_0$ and $\alpha$. (2) Integrate the forward model from $t = 0$ to $t = T$. (3) Compute the cost function $J(q_0, \alpha)$. (4) Integrate the adjoint model backward and compute the gradients. (5) Use an optimization algorithm to find a new estimate of $q_0$ and $\alpha$. (6) If the convergence criteria are satisfied, terminate the procedure; otherwise take the estimated $q_0$ and $\alpha$ as a new guess to repeat (2)–(6).

This procedure, involving the gradient obtained by the conventional or NPE methods, is realized in the next section.

## 4. Numerical experiments

In this section, the estimation of the initial condition and parameter through data assimilation is examined in the time discrete case. A traditional forward time difference scheme is employed to discretize model (3.1).

The discrete form of eq. (2.2) is

$$J_d(q_0, \alpha) = \frac{1}{2}\sum_{k=1}^{K} D_k^2 \Delta t,$$

where $D_k = q_k - q_{\mathrm{obs,k}}$. $q_k$ is the model state, obtained from integration of the discrete model. $q_{\mathrm{obs,k}}$ is the observation. Without considering the observed error and model error, it is evaluated as the analytical solution to eq. (3.1) at $k\Delta t$ with the initial value $q_{\mathrm{obs,0}} = 0.3$. The relevant parameters are $F = 4.0$, $\alpha = -7.0$, $q_c = 0.46$, and $T = 1.0$. Two time resolutions are adopted: $\Delta t = 0.05$ ($K = 20$) and $\Delta t = 0.005$ ($K = 200$). The ranges of the guessed initial condition and parameter value are [0.15, 0.45] and [−8.0, −6.0], respectively.

If $q_k$ is also taken as its analytical value at the $k$ time-step, we obtain the contour of the continuous cost function (as shown in Fig. 1), with a minimum (0.30, −7.0).

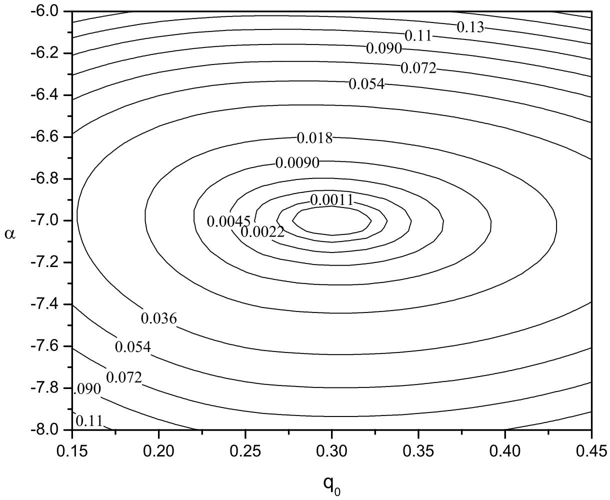


*Fig. 1.* The contours of the continuous cost function.

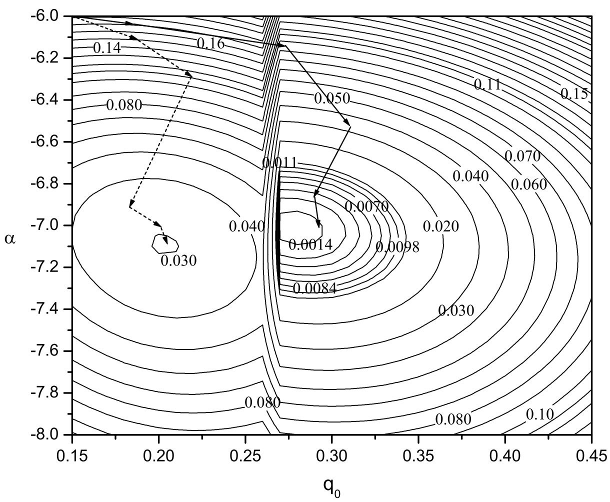


*Fig. 2.* The contours of the discrete cost function with $\Delta t = 0.05$ and minimization procedure for A1 (dashed line) and A2 (solid line) from the first guess (0.15, −6.0).

## 4.1. $\Delta t = 0.05$

When model constraint (3.1) is discretized by the forward time difference scheme, the corresponding cost function has two minima, (0.29, −7.0) and (0.21, −7.1), in the research domain with a discontinuous section (see Fig. 2). The first minimum is the global one and close to the analytical value, while the second is spurious.

By using the conventional method and the NPE method respectively, two minimization approaches can be designed. Method A1 represents the optimization algorithm (M1QN3; Gilbert and Lemarechal, 1989) combined with the conventional method, i.e. the formula of the gradient of $J$ is $J'_d(q_0) = q_0^*$. Method A2 is the optimization algorithm with the NPE method; that is, $J'_d(q_0) = q_0^* + (\alpha/F)q_c q_n^*$, where subscript $n$ in $q_n^*$ means the switch time level. The minimization process from a first guess ($q_{0g} = 0.15$, $\alpha_g = -6.0$) is illustrated in Fig. 2.

The dashed line in Fig. 2 represents the iterative procedure of method A1. The conventional method provides the exact gradient of the discrete cost function (see Mu and Wang, 2003), which is used to determine the next searching direction adopted in the optimization algorithm. Along these descent directions, the iteration converges to the secondary minimum. In method A2 (solid line), the gradient information provided by the NPE method is useful to determine the global descent direction of the discrete cost function, and the search goes through the discontinuous section and then terminates at the global minimum. Comparing these two procedures, we can find their different impacts on the retrieved result in the minimization when there is an artificial minimum on the cost function.

To test the power of A1 and A2 in estimating the optimal values, 441 cases of first guess are used: the first guess of $q_0$ is taken as $0.15 + 0.015 \times i$, $i = 0, \ldots, 20$; the guess of $\alpha$ is $-8.0 + 0.1 \times j$, $j = 0, \ldots, 20$. In Fig. 3, the solid circle represents the first guess ($q_{0g}$, $\alpha_g$), from which the minimization converges to the global minimum; the hollow circle represents the starting point where the iteration terminates at the secondary (or spurious) minimum. It can be seen that the first guess has a tendency to converge to the minimum in the same continuous piece with it, which reflects the local property of the gradient calculated from the conventional method. The cases where the global minimum can be found are about 50.8%. For A2, the percentage increases to 96.4% (see Fig. 4).

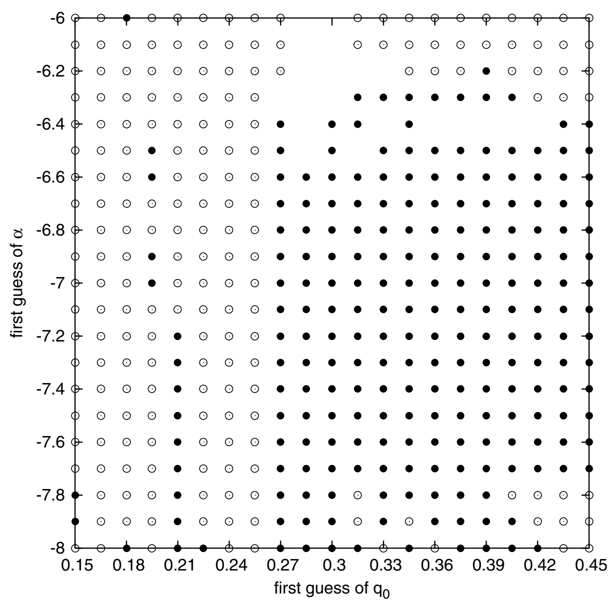


*Fig. 3.* The convergence of the minimization method A1 to the global minimum (solid circle) or secondary minimum (hollow circle) from each first guess grid point.

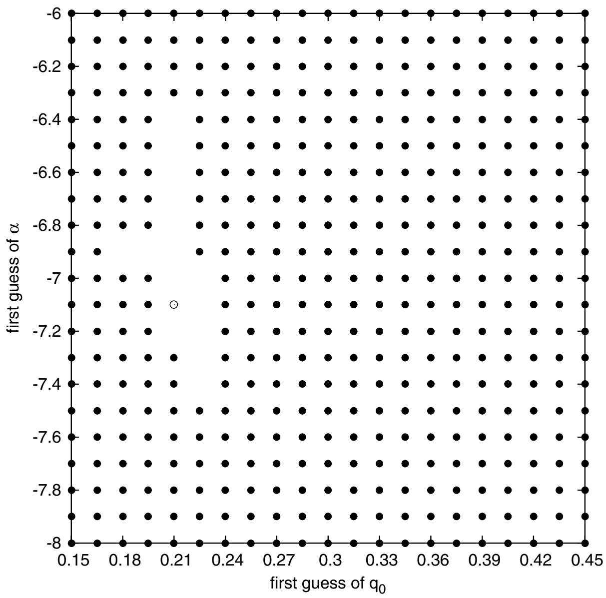


*Fig. 4.* The convergence of the minimization method A2 to the global minimum (solid circle) or secondary minimum (hollow circle) from each first guess grid point.

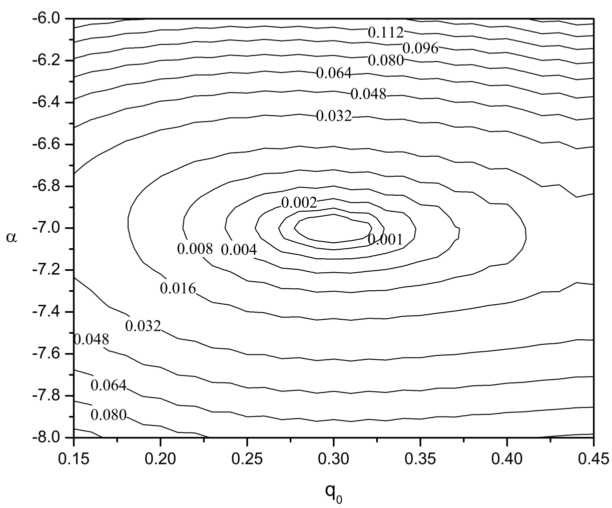


*Fig. 5.* The contours of the discrete cost function with $\Delta t = 0.005$.

### *4.2.* $\Delta t = 0.005$

There is a slight zigzag on the cost function contour (Fig. 5), and two minima, (0.30, −7.0) and (0.29, −7.0), can still be detected even though not shown in the picture. One coincides with the true minimum, and the other close to it. The sum percentage that the first guess taken in the domain converges to (0.30, −7.0) or (0.29, −7.0) is 97.7% for A1, and 99.6% for A2.

From above, we can see that, when the time resolution is coarse, A2 is superior to A1, because the probability of the first guess converging to a spurious minimum is very low (comparing Fig. 4 with Fig. 3). If the time resolution is fine and the minima are close to the true one, there is no big difference between A1 and A2.

## 5. Summary and discussion

The NPE method is applied to the optimal estimation of both model parameter and initial condition by variational data assimilation. This introduces a way to calculate the gradient from the adjoint model. The gradient of the cost function is a combination of the values of this adjoint variable at the initial time and the switch time, which differs from the gradient expression obtained by the conventional method.

Numerical experiments show the impact of gradient information on the recovered initial condition and parameter value. Due to the existence of the jumping ‘on–off’ switch in the constrained model, the traditional discretized cost function may have a zigzag and spurious minimum, while the original cost function is smooth and has only one minimum. In this circumstance, because the gradient calculated by the conventional method is the local gradient, the retrieved initial condition and parameter value depend much on their first guess. If the gradient information is provided through the NPE method, the chance that the optimal initial condition and parameter can be attained is improved greatly, as shown in the cases in this study. The ability of such gradient information in determining the global descent direction of the cost function is illustrated in the minimization procedure (see Fig. 2). In fact, this gradient is the discretized version of the gradient of the original continuous cost function, i.e. in the same form as the coarse-grain gradient (Xu et al., 1998) – but the approach to obtain this gradient is different – whose function is also investigated in Xu and Gao (1999). If the time resolution in discretization is high and the zigzag is not serious on the cost function contour, the global and additional minima are close to the true minimum, and both A1 and A2 have high power in the estimation.

The use of the adjoint technique in the context of physical processes with parametrized discontinuities is a complicated and difficult topic. Because the above result is obtained from an idealized simple model, the application of the NPE method to the real complex model needs further investigation.

## 6. Acknowledgments

The authors would like to thank Dr Qin Xu for his helpful comments, and the encouragement from Dr Xuebin Zhang is also appreciated. This study was supported by the National Nature Science Foundation of China (Grants 4023 3029 and 4022 1503), CAS Key Innovation Direction Project (Grant No. KZCX2-208), and the Innovation Foundation of Institute of Atmospheric Physics, CAS (Grant No. 8-1315).

## References


Fillion, L. and Mahfouf, J.-F. 2000. Coupling of moist-convective and stratiform precipitation processes for variational data assimilation. *Mon. Wea. Rev.* **128**, 109–124.

Gilbert, J. C. and Lemarechal, C. 1989. Some numerical experiments with variable-storage quasi-Newton algorithms. *Math. Programm.* **45**, 407–435.

Le Dimet, F. and Talagrand, O. 1986. Variational algorithms for analysis and assimilation of meteorology observations: theoretical aspects. *Tellus* **38A**, 97–110.

Lu, J. and Hsieh, W. W. 1998. On determining initial conditions and parameters in a simple coupled atmosphere–ocean model by adjoint data assimilation. *Tellus* **50A**, 534–544.

Marecal, V. and Mahfouf, J.-F. 2000. Variational retrieval of temperature and humidity profiles from TRMM precipitation data. *Mon. Wea. Rev.* **128**, 3853–3866.

Mu, M. and Wang, J. F. 2003. A method for adjoint variational data assimilation with physical 'on–off' processes. *J. Atmos. Sci.* **60**, 2010–2018.

Navon, I. M. 1997. Practical and theoretical aspects of adjoint parameter estimation and identifiability in meteorology and oceanography. *Dyn. Atmos. Oceans* **27**, 55–79.

Xu, Q. 1996. Generalized adjoint for physical processes with parametrized discontinuities. Part I: Basic issues and heuristic examples. *J. Atmos. Sci.* **53**, 1123–1142.

Xu, Q. 1998. Comments on "Tangent linear and adjoint of 'on–off' processes and their feasibility for use in four-dimensional variational data assimilation". *Tellus* **50A**, 653–656.

Xu, Q. and Gao, J. D. 1999. Generalized adjoint for physical processes with parametrized discontinuities. Part VI: Minimization problems in multidimensional space. *J. Atmos. Sci.* **56**, 994–1002.

Xu, Q., Gao, J. D. and Gu, W. 1998. Generalized adjoint for physical processes with parametrized discontinuities. Part V: Coarse-grain adjoint and problems in gradient check. *J. Atmos. Sci.* **55**, 2130–2135.

Zhu, Y. Q. and Navon, I. M. 1999. Impact of parameter estimation on the performance of the FSU global spectral model using its full-physics adjoint. *Mon. Wea. Rev.* **127**, 1497–1517.

Zou, X. 1997. Tangent linear and adjoint of 'on–off' processes and their feasibility for use in four-dimensional variational data assimilation. *Tellus* **49A**, 3–31.

Zou, X. 1998. Reply to "Comments on Tangent linear and adjoint of 'on–off' processes and their feasibility for use in four-dimensional variational data assimilation". *Tellus* **50A**, 657–664.

Zou, X., Navon, I. M. and Sela, J. G. 1993. Variational data assimilation with moist threshold processes using the NMC spectral model. *Tellus* **45A**, 370–387.