Discontinuous Forcing Generating Rough Initial Conditions in 4DVAR Data Assimilation

CHUNGU LU AND GERALD L. BROWNING

NOAA/ERL Forecast Systems Laboratory, Boulder, Colorado, and Cooperative Institute for Research in the Atmosphere, Colorado State University, Fort Collins, Colorado

(Manuscript received 23 March 1998, in final form 22 July 1999)

ABSTRACT

The impact of discontinuous model forcing on the initial conditions obtained from 4DVAR data assimilation is studied with mathematic analyses, idealized numerical examples, and more realistic meteorological cases. The results show that a discontinuity in a parameterization, like a model bias, can introduce a systematic error in the assimilated initial fields. However, the most detrimental effect of a model discontinuity is the retention of roughness in the assimilated initial fields, although in some cases the 4DVAR procedure provides some smoothing effect. The obvious consequences of this roughness is that it will introduce spurious modes in the ensuing forecast, and derivatives of the assimilated initial data will be unrealistically large, which can lead to large errors in data analysis. The smoothing effect on the initial conditions with the addition of artificial diffusion to the constraining model is also studied. Possible solutions to the problem of 4DVAR data assimilation with discontinuous model forcing are discussed.

1. Introduction

In recent years, a large number of studies have been devoted to the problem of four-dimensional variational (4DVAR) data assimilation when the constraining equations contain discontinuous forcing terms (Verlinde and Cotton 1993; Bao and Warner 1993; Bao and Kuo 1995; Xu 1996a,b; 1997a,b; Zupanski 1993; Zou et al. 1993; Zou 1997). This problem arises in the implementation of the 4DVAR method for a typical atmospheric model, where the model physical parameterizations usually involve on–off switches. Most of these studies try to determine if the 4DVAR method will still work when the model used as a constraint includes such a switch, especially if the switch is dependent upon the model state variables. From the variational point of view, these concerns are understandable.

Suppose that a forcing term has a jump at a certain model state, for example, $G(u, x, t)$ becomes discontinuous at $u = u_c(x, t)$, where G represents a general forcing, u denotes a model state, the subscript c indicates a critical state when the forcing G has a jump, and x and t are space and time coordinates. The variation of the functional I, which is the combination of the cost function and model constraint (in the Lagrangian form), with respect to the model state variable, $\delta I/\delta u$, would involve a problematic term, $\partial G/\partial u$. Even if this term would not cause a problem in the sense of the general Gateaux variation (Bao and Warner 1993; Zou 1997), there is still the question of whether the solution of the adjoint equation so derived provides a gradient accurate enough to minimize the desired problem.

In this study, instead of concern for the viability of the 4DVAR method in the presence of discontinuous forcing (which has been investigated extensively in the aforementioned studies), an analysis and numerical experiments are devised to examine the impact of such a discontinuity on a 4DVAR solution. Given $u = u(x, t)$, a general forcing $G(u, x, t)$ can be mapped to another function $F(x, t) = G[u(x, t), x, t]$. This treatment can be realized in modeling as the following: A model with the correct functional form of the forcing could be run and the forcing at each time step of the integration saved on disk. Then the model could be run with the forcing for each step being read in from the disk. In the latter case, the forcing behaves exactly as if it were a prescribed function of the independent variables (Browning and Kreiss 1997). Therefore, the discontinuity in the model state associated with the function $G(u, x, t)$ can be treated as a discontinuity in space and time associated with the function $F(x, t)$. Such a treatment formally avoids the problem in the variation, for the forcing $F(x, t)$ drops out of the adjoint equation (the variation of F with respect to state-dependent variable is zero). However, since the discontinuous forcing exists in the forward equation, the integration of the adjoint equation

Corresponding author address: Dr. Chungu Lu, NOAA/ERL R/E/FS, Forecast Systems Laboratory, 325 Broadway, Boulder, CO 80303. E-mail: lu@fsl.noaa.gov

© 2000 American Meteorological Society

Unauthenticated | Downloaded 08/04/21 12:43 PM UTC

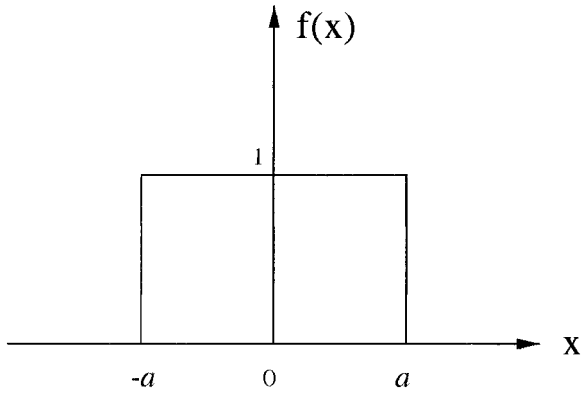

FIG. 1. A step function.

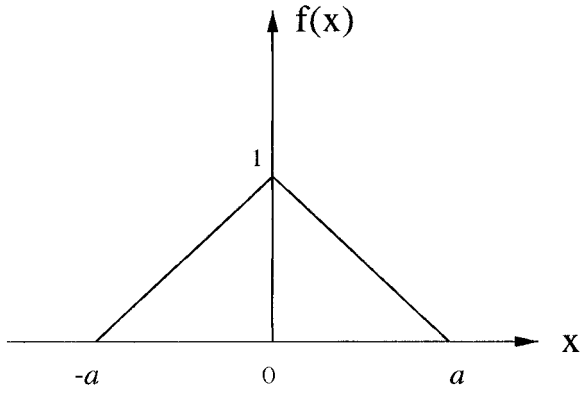

FIG. 2. A triangle pulse function.

will be affected by the discontinuity in two ways: the solution of the forward equation will enter the adjoint equation 1) as an adjoint forcing term and 2) as variable coefficients in the advection terms. Therefore, the discontinuity will still show its impact on the assimilation process. The focus of this study is to examine what kind of discontinuous signature is present in the initial fields obtained from the 4DVAR procedure as a result of a rough forcing function (the exact definition of which will be given in the section that follows). The implications of this signature on the subsequent forecast and possible solutions to the resultant problem are also discussed.

The plan for the paper is as follows. In section 2, the mathematical theory associated with the smoothness of a function is reviewed. This theory helps in understanding and interpreting the theoretical results derived in section 3. The analysis conducted in section 3 is similar to the analysis in Lu and Browning (1998) but is extended from the ordinary differential equation case to the partial differential equation (PDE) case. In section 4, 4DVAR data assimilation experiments are conducted for a simple idealized problem. In section 5, a 4DVAR data assimilation for the barotropic vorticity model is performed. Conclusions are provided in section 6.

2. A mathematical theorem of smoothness revisited

When one is trying to determine the smoothness of a given function, one can examine the Fourier coefficients of the function as a function of wavenumber, that is, the spectrum of the function. In regard to this issue, the following mathematical theorem is relevant (e.g., Gustafsson et al. 1995): Let $f(x)$ be a 2π-periodic function and assume that its pth derivative is a piecewise C^1-function. Then

$$|\hat{f}(k)| \leq \text{constant}/(|k|^{p+1} + 1), \tag{2.1}$$

where k is the wavenumber, x can be any independent variable, $C^1(a, b)$ is the class of functions with one continuous derivative on $a \leq x \leq b$, and the circumflex denotes a spectral variable (and in all following analysis).

To understand the theorem, consider the following two examples. A step function (see Fig. 1),

$$f(x) = \begin{cases} 1, & |x| \leq a, \\ 0, & |x| > a, \end{cases} \tag{2.2a}$$

has zero piecewise derivatives, because the left and right limits at the jump points are not the same; for example, $\lim_{x\to a^+} f(x) \neq \lim_{x\to a^-} f(x)$. As a result, the one-sided, first derivatives at the jump points are not all defined. For example, the left derivative at jump point a is $\lim_{h\to 0^-} [f(a + h) - f(a)]/h = 0$, while the right derivative at jump point a is $\lim_{h\to 0^+} [f(a + h) - f(a)]/h = \infty$. The spectral coefficients for this function can be obtained using the Fourier formula and can be estimated as

$$|\hat{f}(k)| = \left|\frac{\sin ak}{k}\right| \leq \frac{1}{|k|}, \qquad (k \neq 0); \tag{2.2b}$$

that is, the spectral fall-off rate for this step function is on the order of $1/|k|$.

A triangle pulse function (see Fig. 2),

$$f(x) = \begin{cases} 1 - |x|/a, & |x| \leq a \\ 0, & |x| > a, \end{cases} \tag{2.3a}$$

has one piecewise derivative, because it is continuous; for example, $\lim_{x\to 0^+} f(x) = \lim_{x\to 0^-} f(x)$. Therefore its one-sided, first derivatives at the kink points, for example, $\lim_{h\to 0^+} [f(0 + h) - f(0)]/h$ and $\lim_{h\to 0^-} [f(0 + h) - f(0)]/h$, exist. However, its second derivatives do not exist for the same reason as for the step function. Note that the left and right derivatives are $-1/a$ and $1/a$, respectively. Therefore, the derivative at the kink point does not exist, although the one-sided derivatives are definable.

The spectral coefficients for this function satisfy the estimate

$$|\hat{f}(k)| = \left|\frac{4 \sin^2 ak/2}{ak^2}\right| \leq \frac{1}{|k|^2}, \qquad (k \neq 0); \tag{2.3b}$$

Unauthenticated | Downloaded 08/04/21 12:43 PM UTC

that is, the spectral fall-off rate for a triangle pulse function is on the order of $1/|k|^2$.

It is indicated from this theorem that a function would be smoother if its Fourier coefficients were bounded by higher powers of the wavenumber in the denominator, that is, if its spectrum falls off relatively faster. The fact that a Gaussian type of spectral function $\hat{f}(k) = \sqrt{\pi}\exp(-k^2/4)$ converges faster than a function of any inverse power of k indicates an infinitely differentiable function in physical space, in this case known as $f(x) = \exp(-x^2)$.

For a nonperiodic function, Fourier integration needs to be considered to show that the theorem still applies. A heuristic proof of this concept for this case is as follows. Consider a function $f(x)$ that is defined on the entire real line and expandable as a Fourier integral. Then it can be expressed as

$$f(x) = \frac{1}{2\pi}\int_{-\infty}^{\infty} \hat{f}(k)e^{ikx}\, dk.$$

If its pth derivative exists and is expandable as a Fourier integral, then

$$f^{(p)}(x) = \frac{1}{2\pi}\int_{-\infty}^{\infty} (ik)^p \hat{f}(k)e^{ikx}\, dk. \tag{2.4}$$

According to a convergence theorem for infinite integrals, the above integral will converge if

$$|(ik)^p \hat{f}(k)e^{ikx}| \le |k|^p|\hat{f}(k)| < \frac{1}{|k|}, \qquad (k > 1), \tag{2.5}$$

which is essentially the statement (2.1).

The following definitions are used throughout this paper. The terminology ''smooth function'' is used for a function that possesses at least several derivatives; otherwise the function is referred to as a ''rough function.''

In a numerical model with a fixed resolution, the impact of a discretized rough or smooth forcing on the solution is different. It is probably obvious that one needs almost an infinite number of grid points to resolve a discontinuous forcing. Revisiting the first example, one can see that spectrally speaking, a sufficiently large number of waves must be used in the numerical model in order to capture most of the slowly decaying spectrum of a step function. This implies that if there exists a rough forcing function in the model, numerical accuracy of the solution may not be achievable with a reasonable resolution (limited numbers of waves). On the other hand, a smooth forcing function can be resolved with a reasonable number of waves, and the accuracy of the numerical solution can be maintained.

3. Analysis

Consider minimizing the cost function

$$I = \iint w(x, t)|u - u_{\text{obs}}|^2\, dx\, dt, \tag{3.1}$$

subject to

$$u_t + au_x = F(x, t), \tag{3.2}$$

where u and u_{obs} are the model and observational state variables, respectively, a is the constant advection speed; and $w(x, t)$ is a weighting function. In practice (e.g., for a discrete analysis), this weighting function is related to an observational or model background error covariance matrix. To simplify the presentation in the subsequent analyses, let $w(x, t)$ be 1 (e.g., Talagrand and Courtier 1987; Lu and Browning 1998). This analysis can be generalized to a set of hyperbolic PDEs.

The model physical forcing, $F(x, t)$, includes the true physical process (assumed smooth) and a discontinuous part due to the parameterization. Although a general forcing can also be a function of the dependent variable, for example, $F = F(u, x, t)$, such a general forcing function can still be considered as only a function of (x, t), given $u = u(x, t)$. The focus of the current study is to examine the possible impact of a discontinuous forcing on the assimilated initial condition. For this purpose, the physical forcing and its discontinuity are assumed to be a function of space and time only. Because F is independent of u, the variation of the constraint equation with respect to u is straightforward, that is, without the complication of the term $\partial F/\partial u$. Therefore, this simplification avoids the question as to whether the 4DVAR method would work and how it would work when the assimilated model possesses discontinuous forcing terms, which has been extensively studied in Bao and Warner (1993), Xu (1996a), and Zou (1997). However, because the discontinuous forcing affects the solution of the forward equation (3.2), and therefore the adjoint equation, an effect of the discontinuity remains in the assimilated initial conditions obtained by the 4DVAR approach.

Consider the Fourier transformation

$$u(x, t) = \frac{1}{2\pi}\int_{-\infty}^{\infty} \hat{u}(k, t)e^{ikx}\, dk \tag{3.3a}$$

and its inverse transformation

$$\hat{u}(k, t) = \frac{1}{2\pi}\int_{-\infty}^{\infty} u(x, t)e^{-ikx}\, dx. \tag{3.3b}$$

Using (3.3a) and (3.3b), one can transform (3.2) to Fourier space as

$$\hat{u}_t + iak\hat{u} = \hat{F}, \tag{3.4}$$

which is readily solved in the form

$$\hat{u} = \hat{u}(k, 0)e^{-iakt} + \int_0^t \hat{F}(k, \tilde{t})e^{iak(\tilde{t}-t)}\, d\tilde{t}, \tag{3.5}$$

where $\hat{u}(k, 0)$ is the initial condition corresponding to time $t = 0$, and t can be either positive or negative. Assuming that the true atmospheric state is completely observed and there is no observational error, one can write down the observed state in the form

Unauthenticated | Downloaded 08/04/21 12:43 PM UTC

$$\hat{u}_{\text{obs}} = \hat{u}_{\text{true}}(k, 0)e^{-iakt} + \int_0^t \hat{F}_{\text{true}}(k, \tilde{t})e^{iak(\tilde{t}-t)}\, d\tilde{t}, \quad (3.6)$$

where $\hat{u}_{\text{true}}(k, 0)$ and $\hat{F}_{\text{true}}$ denote the true initial condition and the true physical forcing. Assume that the true forcing is a smooth function of both time and space. Therefore, $\delta\hat{F} = \hat{F}_{\text{true}} - \hat{F}$ contains the discontinuity artifact in the parameterizations.

On substituting (3.5) and (3.6) into (3.1), one can rewrite the cost function as

$$I = \iint |u - u_{\text{obs}}|^2\, dx\, dt = \iint |\hat{u} - \hat{u}_{\text{obs}}|^2\, dk\, dt = \iint \left| [\hat{u}(k, 0) - \hat{u}_{\text{true}}(k, 0)]e^{-iakt} - \int \delta\hat{F}e^{iak(\tilde{t}-t)}\, d\tilde{t} \right|^2 dk\, dt,$$

where the second equality is obtained by Parseval's relation.

A proper initial condition, $\hat{u}(k, 0)$, that minimizes the above cost function can be determined by taking a derivative of the functional I with respect to $\hat{u}(k, 0)$, setting the result to zero, and solving for $\hat{u}(k, 0)$. However, one must be careful about the procedure, because all spectral variables are complex quantities. Using the formula $|z|^2 = z\overline{z}$, where $z = z_{\text{R}} + iz_{\text{I}}$ is a complex variable, $\overline{z} = z_{\text{R}} - iz_{\text{I}}$ is the conjugate of z, and z can be $\hat{u}(k, 0)$, $\hat{u}_{\text{true}}(k, 0)$, or $\delta\hat{F}$, one can write the cost function I in terms of real and imaginary parts. The minimization requires

$$\frac{\partial I}{\partial \hat{u}_R(k, 0)} = 0, \qquad \frac{\partial I}{\partial \hat{u}_I(k, 0)} = 0,$$

which give two conditions,

$$e^{-iakt}\overline{A} + e^{iakt}A = 0, \qquad e^{-iakt}\overline{A} - e^{iakt}A = 0,$$

where $A = [\hat{u}(k, 0) - \hat{u}_{\text{true}}(k, 0)]e^{-iakt} - \int \delta\hat{F}e^{iak(\tilde{t}-t)}\, d\tilde{t}$. Combining the two conditions, and solving for $\hat{u}(k, 0)$, one obtains

$$\hat{u}(k, 0) = \hat{u}_{\text{true}}(k, 0) + \int_0^t \delta\hat{F}(k, \tilde{t})e^{iak\tilde{t}}\, d\tilde{t}. \quad (3.7)$$

This result indicates that when there is a discontinuity in a state-independent forcing, the 4DVAR-derived initial condition will converge to the true initial condition plus a term generated by the error in the forcing.

The integral and the exponential factor combined can sometimes reduce the effect of a discontinuous forcing error. To see this, first consider a case in which a discontinuity occurs in time, but forcing in space is a continuous function and can be separated from the forcing in time; that is, $\delta\hat{F}(k, \tilde{t}) = \hat{S}(k)D(\tilde{t})$. In this case, (3.7) becomes

$$\hat{u}(k, 0) = \hat{u}_{\text{true}}(k, 0) + \hat{S}(k) \int_0^t D(\tilde{t})e^{iak\tilde{t}}\, d\tilde{t}. \quad (3.8)$$

Assuming that $D(\tilde{t})$ is a step function in time and has a jump between t_1 and t_2 ($\Delta t = t_2 - t_1 \in [\tau, 0]$, where τ is the beginning assimilation time and 0 is the ending assimilation time, and $\tau < 0$), one obtains

$$\hat{u}(k, 0) = \hat{u}_{\text{true}}(k, 0) + (iak)^{-1}\hat{S}(k)(e^{iakt_1} - e^{iakt_2}), \quad (3.9)$$

which can be bounded as

$$|\hat{u}(k, 0)| \le |\hat{u}_{\text{true}}(k, 0)| + 2\left|\frac{\hat{S}(k)}{ak}\right|. \quad (3.10)$$

Note that $\hat{S}(k)$ is a smooth function. According to the mathematical theorem reviewed in section 2, the smoothness of $\hat{S}(k)$ is further improved when divided by k in (3.10). This result indicates that 1) the initial field derived by the 4DVAR method is smooth when the discontinuity in model forcing occurs only in time, and 2) the time discontinuity, like a model bias, will introduce a systematic error in the assimilated initial field [the addition of the second term on the right-hand side of (3.10)].

Next consider the case when a discontinuity occurs in space. Assuming that the forcing in temporal part is continuous and separable from the forcing in spatial part, one can write (3.7) in the form

$$\hat{u}(k, 0) = \hat{u}_{\text{true}}(k, 0) + \hat{D}(k) \int_0^t S(\tilde{t})e^{iak\tilde{t}}\, d\tilde{t}, \quad (3.11)$$

where $\hat{D}(k)$ denotes a discontinuous function in space, and $S(\tilde{t})$ denotes a smooth function in time. Depending upon the type of S as a function of time, the impact of the discontinuous forcing in space on the initial field can vary. One of the simple cases that can be analyzed is when $S(\tilde{t}) = 1$, that is, when the forcing term is time independent and discontinuous in space. Then the integral in (3.11) can be evaluated from the beginning assimilation time τ to the ending assimilation time 0, which gives

$$\hat{u}(k, 0) = \hat{u}_{\text{true}}(k, 0) + (iak)^{-1}\hat{D}(k)(e^{iak\tau} - 1), \quad (3.12)$$

and can be bounded as

$$|\hat{u}(k, 0)| \le |\hat{u}_{\text{true}}(k, 0)| + 2\left|\frac{\hat{D}(k)}{ak}\right|. \quad (3.13)$$

If the discontinuous forcing is a step function, then $\hat{D}(k)$

Unauthenticated | Downloaded 08/04/21 12:43 PM UTC

will decrease like $1/|k|$, as in the example discussed in the previous section. It is clear from (3.13) that the 4DVAR-derived initial condition will converge to the true initial condition with a supposition of a rough function whose spectrum falls off like $1/|k|^2$. Thus, in this case some degree of smoothing can be achieved with the 4DVAR minimization procedure. However, only one order of smoothing is achieved (i.e., the spectral falloff rate is improved by only $1/|k|$) with the 4DVAR method. If $S(\tilde{t})$ is a complicated function of time, smoothing may not be achievable. The simplest example is when $S(\tilde{t}) = e^{-iak\tilde{t}}$, in which case the S term cancels the $e^{iak\tilde{t}}$ term in (3.11), resulting in no smoothing of the discontinuous forcing $\hat{D}(k)$.

Summarizing the above analyses, one can see the following results. Discontinuities in the model forcing due to parameterizations, like a model bias, will generally introduce errors in the initial fields when the 4DVAR procedure is performed. The 4DVAR method has some smoothing effect on a discontinuous forcing. For discontinuous forcing in time, the assimilated data will not show any discontinuous signature other than that a systematic error is incorporated. In the case of a spatial discontinuity in the model forcing, one order of smoothing is gained when the forcing is a function of space only, and smoothing may or may not be achievable when the forcing is a complicated function of both space and time. The implication of the above results is clear: if a step function type of discontinuity exists in a forecast model (discontinuous in space, in particular), the 4DVAR-assimilated initial conditions will not have sufficient smoothness to ensure a slowly varying solution of the prediction model. The spurious modes generated by the nonsmooth initial condition can contaminate the physical solutions. This issue has been studied in association with the bounded derivative principle (Kreiss 1980; Browning and Kreiss 1994). Furthermore, because of the roughness in the initial field, the derivatives of the initial field (e.g., the vorticity or divergence field derived from the initial wind field) may not exist. Therefore, if one tries to obtain a derivative field based on this rough initial data, one may get a nonphysical result.

In some atmospheric models, diffusion may help to smooth the roughness in the 4DVAR-derived initial field(s). However, the form and value of the diffusion coefficients for various physical processes or model variables are not well known. For example, the choice of the set of diffusion coefficients for water vapor, cloud water, cloud ice, etc., in microphysics equations is quite empirical. Furthermore, improper use of diffusion to smooth the roughness in numerical solutions often causes a decrease of numerical accuracy. On the basis of these arguments, it is suggested from this analysis that the best solution for the 4DVAR data assimilation with a model subject to a discontinuous forcing is to smooth the discontinuities in the physical parameterizations, and that the smoothing procedure needs also to take into consideration that the smoothed model forcing needs to be close to the physical forcing so that its impact on forecast accuracy may be minimized.

4. Numerical solutions

To verify the results from the previous analysis, we conducted numerical 4DVAR data assimilations. In this section, the numerical solution is sought for the problem of minimizing (3.1), subject to (3.2), via the adjoint equation

$$U_t + aU_x = (u - u_{\text{obs}}), \tag{4.1}$$

where U is the adjoint variable (the Lagrangian multiplier), and the nondimensional value $a = 1$ is chosen. The adjoint equation can be derived by taking the variation of the functional, defined by the combination of (3.1) and (3.2), with respect to u.

In the first experiment, the model forcing in (3.2) is specified as $F(x, t) = S(x)D(t)$, where $S(x)$ is a smooth function of space (see Fig. 3a) and $D(t)$ is a step function in time (see Fig. 3b), to mimic a switch that is turned on and off in a physical parameterization. The true forcing is a continuous function both in time and space with steep variations (see Figs. 3a and 3b). A real valued function in C^∞ (infinitely differentiable) is used to represent the steep changes of the step function. This function is expressed as

$$f(x) = \alpha \frac{C(x)C(x_s - x)}{[C(x) + C(x_s - x)]^2}, \tag{4.2}$$

where $C(x)$ is a positive, continuous function, defined as

$$C(x) = \begin{cases} 0, & x \le 0, \\ e^{-\beta/x}, & x > 0, \end{cases} \tag{4.3}$$

where the constants $\alpha = 4$ and $\beta = 0.5$ are chosen, and x_s is used to shift the origin. This function is used both in space and time with x_s chosen accordingly. The observational field is obtained by integrating (3.2) numerically subject to this true forcing.

Integrating (3.2) forward in time and (4.1) backward in time, one can find the gradient of the cost function. Following this gradient, one can minimize the cost function using some kind of minimization procedure. In this study, the steepest descent method is used. The whole procedure is repeated until a minimum is achieved; for details, see Lu and Browning (1998). All experiments starts with a initial guess of 0. The resolution used in this case is 100 points for domains in plots. It takes only three cycles (a cycle consists of one sweep of forward integration, one sweep of backward integration, another sweep of forward integration, and iteration of the steepest descent procedure) for the solution to reach the minimum.

Figure 3c is a plot of the true initial and assimilated initial conditions. One can see that both curves are smooth, apart from the fact that the assimilated initial

Unauthenticated | Downloaded 08/04/21 12:43 PM UTC

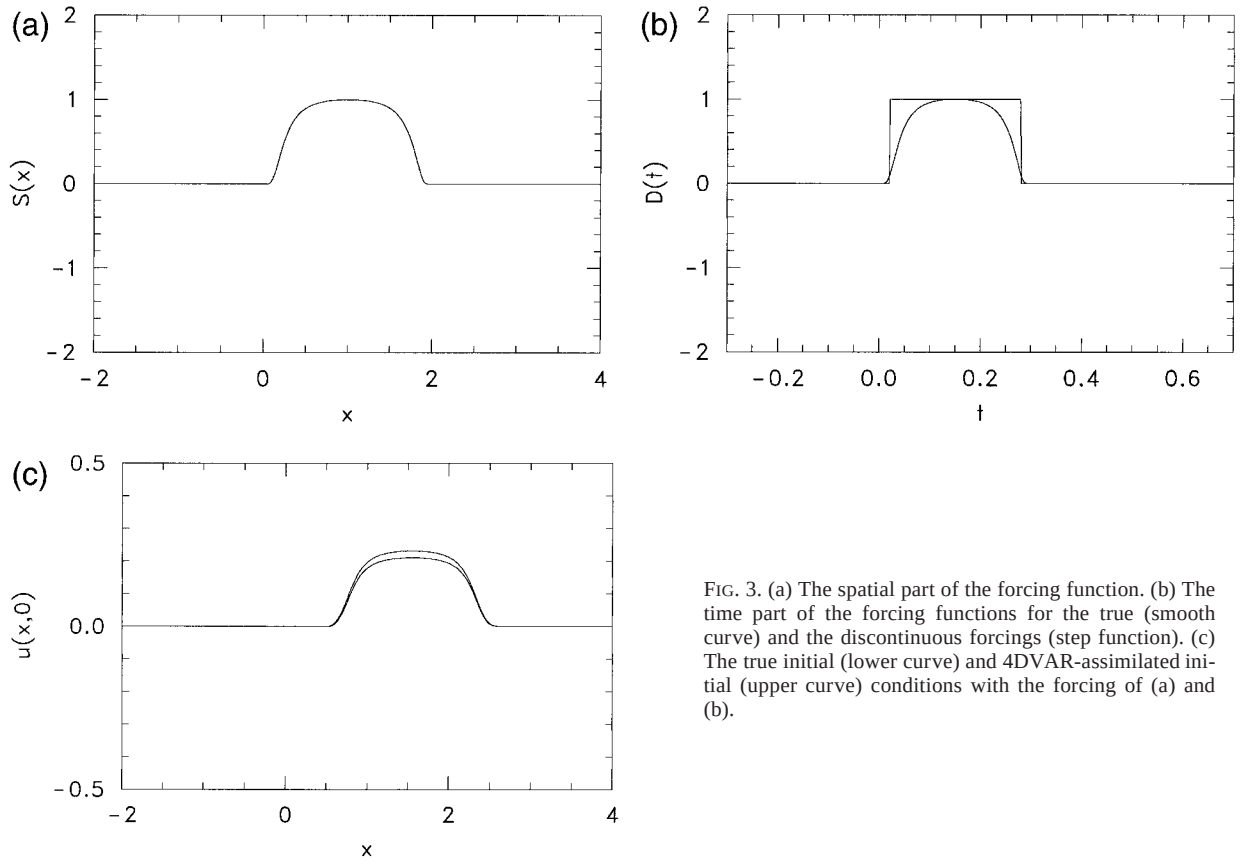

FIG. 3. (a) The spatial part of the forcing function. (b) The time part of the forcing functions for the true (smooth curve) and the discontinuous forcings (step function). (c) The true initial (lower curve) and 4DVAR-assimilated initial (upper curve) conditions with the forcing of (a) and (b).

condition is systematically different from the true condition. The relative L_2 error between the true and assimilated initial conditions is about 14.1%.

In the second experiment, consider the case that the time-independent forcing term has a discontinuity in space, that is, $F(x, t) = D(x)$. The step function in Fig. 4a represents this forcing term, and the smooth curve [given by Eq. (4.2)] is assumed to be the true forcing. Again, 100 points are used for the computational domain. The solution reached the minimum after 91 cycles of integration, indicating that the roughness in the forward solution substantially slows down the convergence process. The relative L_2 error between the true and assimilated initial condition is 6.96%. The assimilated initial condition is composed of a smooth part (close to the true initial condition) and a rough part (a kink) re-

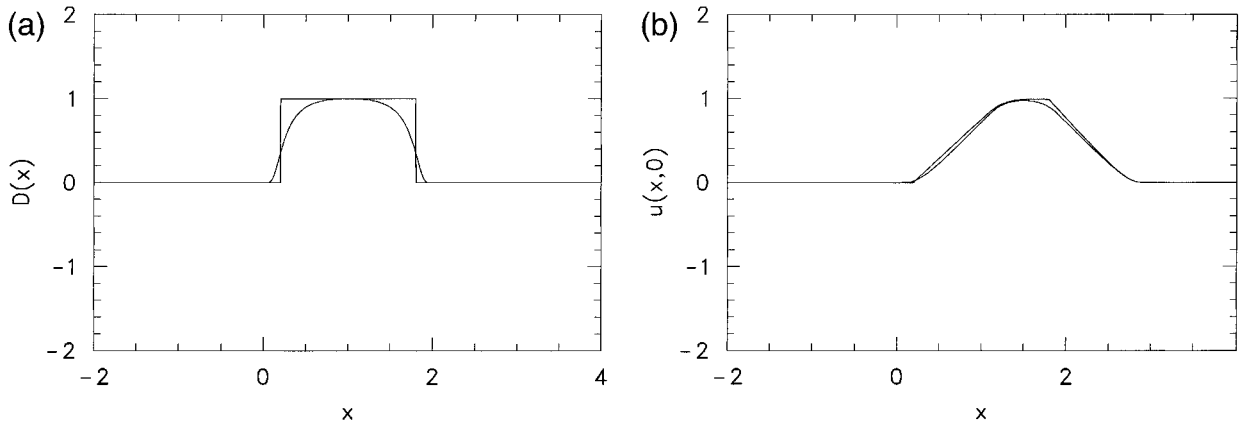

FIG. 4. (a) The forcing functions for the true (smooth curve) and the discontinuous forcings (step function). (b) The true initial (smooth curve) and 4DVAR-assimilated initial (rough curve) conditions with the forcing of (a).

Unauthenticated | Downloaded 08/04/21 12:43 PM UTC

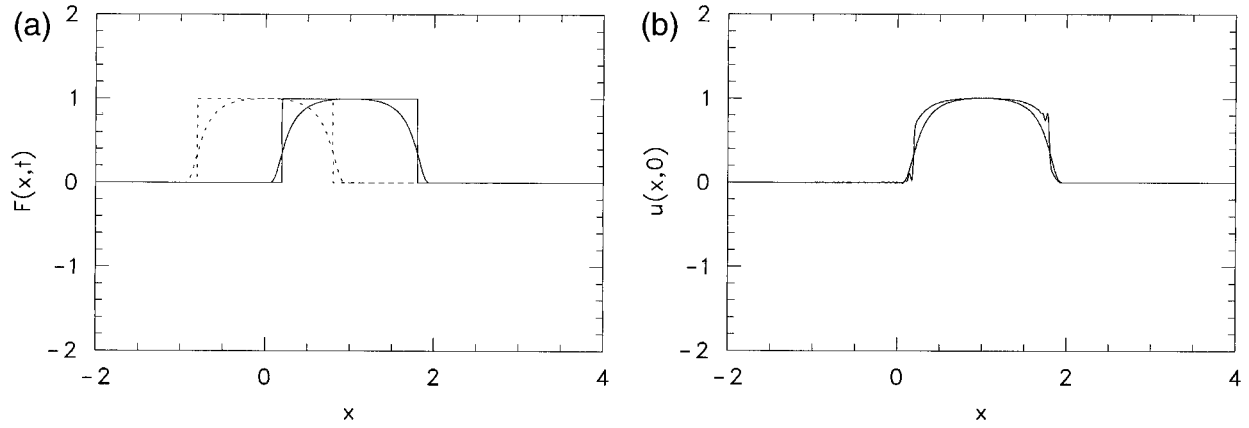

FIG. 5. (a) The translating forcing functions for the true (smooth curve) and the discontinuous forcings (step function). (b) The true initial (smooth curve) and 4DVAR-assimilated initial (rough curve) conditions with the forcing of (a).

flecting the influence of the discontinuous forcing (Fig. 4b). This confirms the analysis that 4DVAR data assimilation provides a smoothing effect for a time-independent discontinuous forcing. However, only one order of smoothing is achievable, and this is not sufficient to provide a smooth initial condition for a forecast.

The next case considers a forcing as a function of both time and space, with discontinuities in space and smoothness in time. In particular, the forcing is the translating function in time, that is, $F(x, t) = F(x - at)$, shown in Fig. 5a. The translation speed a is chosen to be the same as the advection speed. As the analysis predicted in section 3, smoothing is not obtainable in this case. Figure 5b shows that the 4DVAR-assimilated initial condition has discontinuities in comparison with the true initial condition. Two spikelike features with Gibb's ripples are present at locations where the forcing function is discontinuous. For this case, the 4DVAR procedure did not have a smoothing effect on forward model solutions at all, and the convergence took 360 cycles to complete.

All these experimental results agree with the analyses conducted in the previous section. In summary, a discontinuous forcing will introduce systematic errors in the 4DVAR-assimilated initial field(s). If the discontinuity occurs in time in the forcing term, the 4DVAR procedure gives a smooth initial field. Thus, a time discontinuous forcing exerts less impact on the 4DVAR data assimilation as far as roughness in the initial condition is concerned. If the discontinuity occurs in space in the forcing term, the 4DVAR procedure will most likely generate a rough initial field, although some smoothing effect can be obtained in certain cases. This roughness in the 4DVAR-assimilated initial conditions has a negative implication on the ensuing forecasts and data analysis, as discussed previously.

In some 4DVAR data assimilations with realistic atmospheric models, the impact of roughness in the forcing on the forecast may not be significantly seen, for example, in Zou et al. (1993) and Zou (1997). One of the possible explanations is that these models include large diffusion terms. The diffusion can add a smoothing effect on the assimilated fields. To see the smoothing effect by diffusion, we conducted experiments by adding a diffusion term to (3.2) and the corresponding term to the adjoint equation (4.1). In this case, the forward and adjoint equations become

$$u_t + au_x = F(x, t) + \nu_1 u_{xx} - \nu_2 u_{xxxx} \quad \text{and} \quad (4.4)$$

$$U_t + aU_x = (u - u_{\text{obs}}) - \nu_1 U_{xx} + \nu_2 U_{xxxx}, \quad (4.5)$$

respectively, where ν_1 is the diffusion coefficient for the second-order diffusion term, and ν_2 is the diffusion coefficient for the fourth-order diffusion term. According to the The Pennsylvania State University–National Center for Atmospheric Research Mesoscale Model version 5 model documentation (Grell et al. 1995), $\nu_2 = (\Delta x)^2 \nu_1$, where Δx is horizontal grid spacing. Both the second- and fourth-order diffusions are added to make sure that high wavenumber noise is sufficiently damped.

Figures 6a and 6b show the true initial conditions and the assimilated initial conditions for the cases of the translating, spatial discontinuous forcing (the same as the case presented in Fig. 5, but with diffusion). Figure 6a is for the nondimensional $\nu_1 = 10^{-4}$, and Fig. 6b is for $\nu_2 = 10^{-3}$ [which is the maximum dissipation that can be used in this particular case when Δt is determined by the Courant–Friedrichs–Levy (CFL) limit for advection]. One can see that by choosing a sufficiently large ν_1, diffusion terms are capable of smoothing out the roughness in the assimilated field(s). The numerical accuracy in these two cases is slightly improved. The relative L_2 errors are 11.42% and 10.92%, respectively, for $\nu_1 = 10^{-4}$ and $\nu_2 = 10^{-3}$, compared with 12.20% for the case when there is no diffusion term. The addition of diffusion in the model did not help the 4DVAR convergence process significantly. For the above two cases, 314 and 283 cycles were required, respectively, to reach the minima.

Unauthenticated | Downloaded 08/04/21 12:43 PM UTC

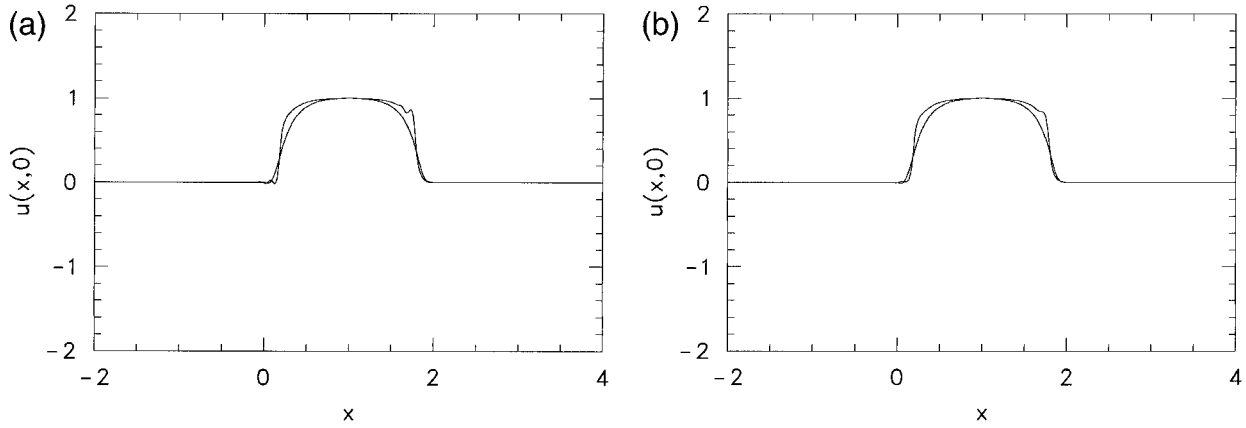

FIG. 6. (a) Same as Fig. 5b, except that diffusions are added to the assimilated model, with the nondimensional diffusion coefficient $\nu_1 = 10^{-4}$. (b) Same as (a), except for the nondimensional diffusion coefficient $\nu_1 = 10^{-3}$.

5. Experiments with the vorticity model

To put the above discussions into a more realistic meteorological context, consider 4DVAR data assimilation with a barotropic vorticity equation forced by a spatially discontinuous energy source. This problem can be viewed as a parameterized physical process that gives rise to a spatially discontinuous heat source, which forces a divergence–convergence field. The vorticity field feels the discontinuous forcing through this divergence–convergence field.

The forward vorticity equation is

$$\zeta_t + J(\Delta^{-1}\zeta, \zeta + f) = F(x, y, t), \tag{5.1}$$

where J is the Jacobian operator, Δ^{-1} is the inverse of the two-dimensional Laplacian operator [following Talagrand and Courtier (1987)], $F(x, y, t)$ denotes the physical forcing term (divergence–convergence), and other notations are standard in meteorology. For simplicity, F is considered as a function of space and has discontinuities only in the x direction. The mathematical expression for this forcing term is

$$F(x, y) = -4r_e^{-2}\Psi_0 f(x) e^{-(y-y_0)^2/r_e^2}, \tag{5.2}$$

where $\Psi_0 = 10^7$ m^2 s^{-1} is the characteristic value of the streamfunction, $r_e = 350$ km is the e-folding radius, $f(x)$ is a step function with a step width of 1000 km (centered at x_0), and (x_0, y_0) is located at the center of the computational domain. Figure 7a is the plot of this forcing function with the abscissa for x and ordinate for y. One can see that the forcing function has discontinuous jumps in the x direction but falls off smoothly (exponentially) in the y direction.

The true forcing (shown in Fig. 7b) is a spatially continuous function with steep variations in the x direction. The expression for this function is the same as (5.2), except that $f(x)$ takes the form of (4.2). The true observations are obtained by integrating the vorticity equation (5.1) subject to the true forcing. Starting with an initial circular ''weather system,'' the integration of the vorticity model to the end of the assimilation period (12 h) will generate a true initial field for the ensuing forecasts. Figure 7c shows this true initial vorticity field.

The adjoint equation can be written (see Lu and Browning 1997) as

$$\zeta'_t + J(\Delta^{-1}\zeta, \zeta') - \Delta^{-1}J(\zeta + f, \zeta') = (\zeta - \zeta_{\text{obs}}), \tag{5.3}$$

where ζ' is the adjoint vorticity (the Lagrangian multiplier). By integrating (5.1) subject to (5.2) forward in time, and (5.3) backward in time, one can obtain the gradient of the cost function. Following this gradient, one tries to minimize the difference between the model and observational vorticity fields. This procedure produces an assimilated initial vorticity field shown in Fig. 7d. Note that the vorticity equation is nonlinear, but that the adjoint equation is linear with variable coefficients. The discontinuous effect will now also enter the assimilation procedure through these coefficients. Because of this effect, the convergence process takes much more time than that for the case of the linear advection equation in section 4. In this particular experiment, it took 550 cycles to reach the minimum.

In comparing Figs. 7c and 7d, one can see that when there is a discontinuous forcing in the assimilation model, the 4DVAR procedure produces an initial condition that is a combination of the true initial field and a component from the discontinuous artifact in the forcing. The roughness in the assimilated initial field is less than in the forcing, but only one order of smoothness is gained. It is clearly shown in the model experiment that with a step function type discontinuity in the model, the 4DVAR-assimilated initial condition will inevitably preserve some roughness in it. When a diffusion term is added, this roughness may be reduced. Figure 7e is a plot of the assimilated initial vorticity field, which is the same as the case shown in Fig. 7d, except that a second-order diffusion $\nu\nabla^2\zeta$ is added to the forward and adjoint equations. The diffusion coefficient $\nu = 10^5$ m^2

Unauthenticated | Downloaded 08/04/21 12:43 PM UTC

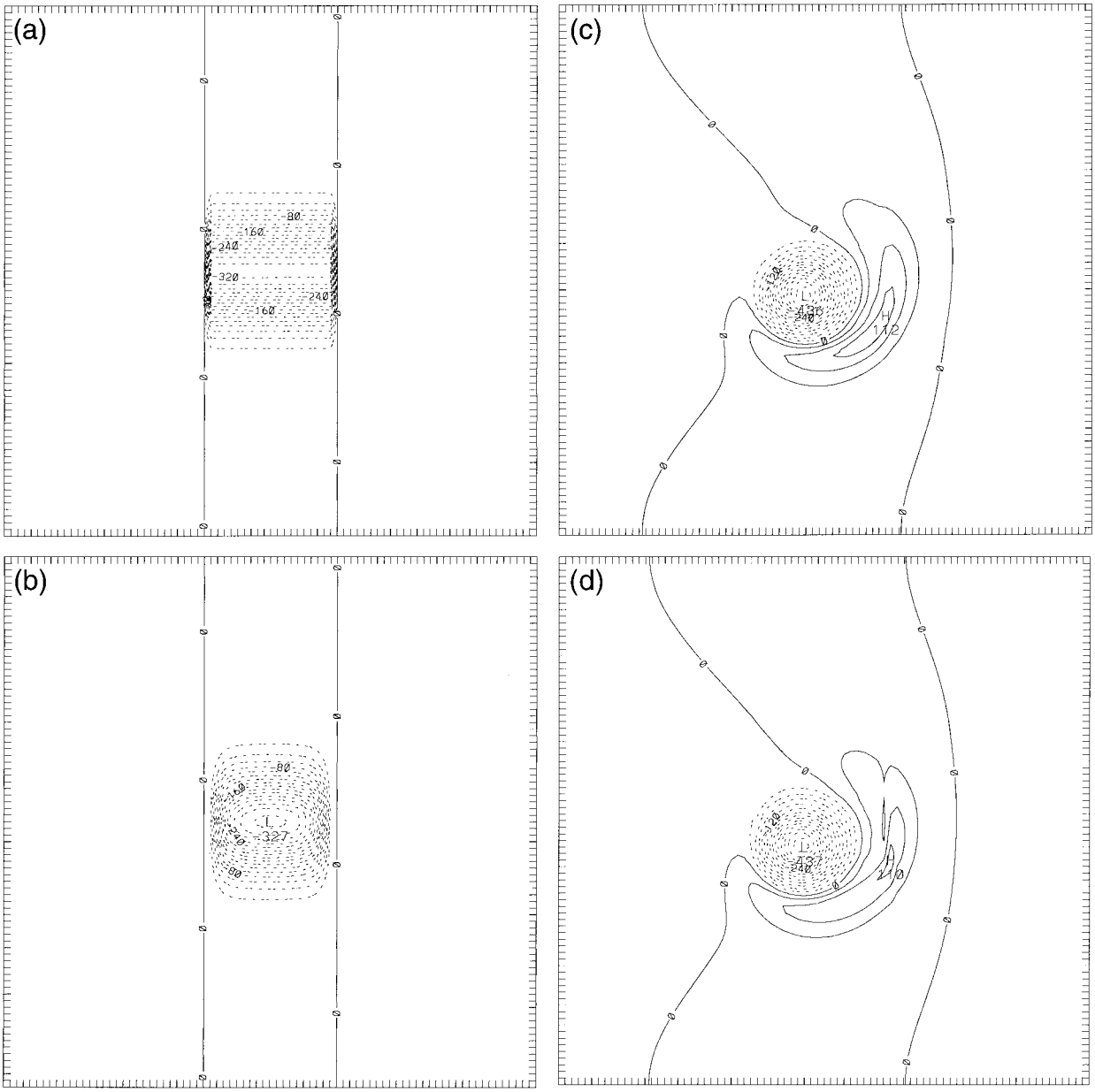

FIG. 7. (a) The parameterized forcing function. The abscissa is for x, and the ordinate is for y. Contours from -0.32×10^{-8} to 0.0 at 0.2×10^{-9} intervals, label scaled by 0.1×10^{12}. (b) The true forcing function. Contours and scaling as in (a). (c) The true initial vorticity field. Contours from -0.4×10^{-3} to 0.9×10^{-4} at 0.3×10^{-4} intervals, labels scaled by 0.1×10^{7}, PT(3.3) = 0.2211×10^{-5}. (d) The assimilated initial vorticity field. Contours and scaling as in (c). PT(3.3) = $0.249\,55 \times 10^{-5}$. (e) The assimilated initial vorticity filed with a diffusion

s^{-1} is used. It is seen from Fig. 7e that the diffusion smooths most of roughness that is present in Fig. 7d. However, in comparison with the true initial vorticity (Fig. 7c), one can see that the diffusion has decreased the numerical accuracy (brought in about a 10% relative error).

6. Concluding remarks

The impact of discontinuous model forcing on 4DVAR data assimilation is studied with mathematic analyses, idealized numerical examples, and more realistic meteorological cases. The commonly used on–off switches in many physical parameterizations are represented as a discontinuous forcing in space and time. Such a treatment formally avoids the problem of the existence of derivatives when a variational procedure is carried out, but the impact of a discontinuity remains in the problem.

A variational analysis is conducted for a forced partial differential equation. With the aid of a mathematical theorem on smoothness of a function, the analyzed re-

Unauthenticated | Downloaded 08/04/21 12:43 PM UTC

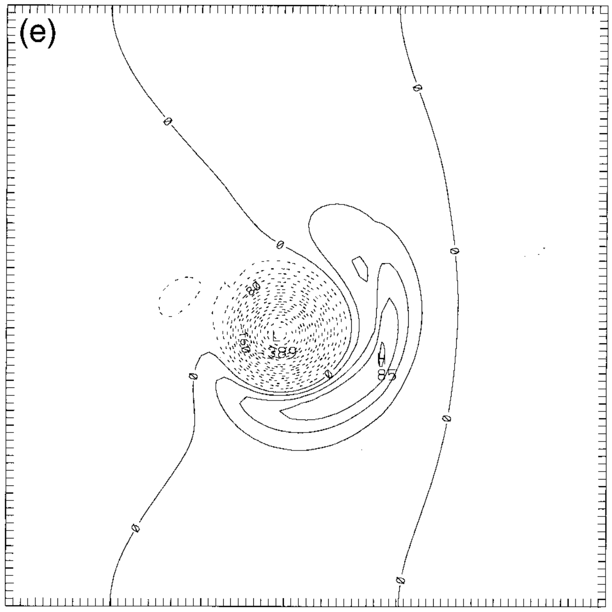

FIG. 7. (*Continued*) term. The diffusion coefficient $\nu = 10^5$ m^2 s^{-1} is used. Contours and scaling as in (c). PT(3.3) = 0.2548×10^{-5}.

sults can be understood, and implications of these results are drawn for the 4DVAR data assimilation with on–off switches in model physical parameterizations. The results show that a discontinuity in the model forcing, like other model bias, can introduce a systematic error in the assimilated initial fields. However, the most detrimental effect of a model discontinuity is the retention of roughness in the assimilated initial fields. It is suggested from the analyses that when a discontinuity occurs only in time in the model forcing, the initial field will be smooth. However, when a discontinuity occurs in space in the model forcing, the 4DVAR-derived initial field will most likely retain some roughness.

The possible consequences of roughness in the assimilated data are the following. First, any derivative field based on these data may not be usable. For example, if one is trying to obtain the vorticity or divergence from the 4DVAR-derived initial wind fields, the roughness can produce large, nonphysical derivatives that have nothing to do with the true vorticity or divergence. Second, forecasts based on this rough initial data will most likely generate artificial, fast waves that may contaminate the true solutions.

Numerical experiments show that use of a large diffusion in the model can provide additional smoothing that can reduce the roughness in the 4DVAR-assimilated initial data. This may explain why there have not been significant problems with the performance of the 4DVAR procedure with some atmospheric models, even though these models have on–off switches in their physical packages. However, due to the lack of understanding of diffusion in various physical processes, the choice of a diffusion coefficient is very empirical. Furthermore, improper use of diffusion to smooth the roughness in numerical solutions often causes a decrease of numerical accuracy (e.g., the vorticity model experiment in section 5). On the basis of these arguments, it is suggested that the best solution for 4DVAR data assimilation with a model subject to a discontinuous forcing is to smooth the discontinuities in the physical parameterizations. This idea has also been proposed in Verlinde and Cotton (1993), Browning and Kreiss (1994), Zupanski (1993), Zupanski and Mesinger (1995), and Tsuyuki (1996).

It should also be pointed out that in practice, it may be possible to reduce the impact of discontinuous model forcing on assimilated solutions by putting more weight on observational data. However, the procedure requires prior knowledge of the model forcing errors and the impact of these errors on the assimilated solutions. The present study tries to gain some understanding of these issues before the weighting procedure.

Acknowledgments. Thanks to Dr. Alexander MacDonald for his support throughout this study. Thanks to three anonymous reviewers for their constructive comments and to Dr. Joseph Tribbia for his excellent editorship. Thanks also to Dr. Yuanfu Xie and Christy Sweet, who carefully conducted NOAA in-house review and edit, respectively.

REFERENCES

Bao, J.-W., and T. T. Warner, 1993: Treatment of on/off switches in the adjoint method: FDDA experiments with a simple model. *Tellus,* **45A,** 525–538.

——, and Y.-H. Kuo, 1995: On–off switches in the adjoint method: Step functions. *Mon. Wea. Rev.,* **123,** 1589–1594.

Browning, G. L., and H.-O. Kreiss, 1994: The impact of rough forcing on systems with multiple time scales. *J. Atmos. Sci.,* **51,** 369–383.

——, and ——, 1997: The role of gravity waves in slowly varying in time mesoscale motions. *J. Atmos. Sci.,* **54,** 1166–1184.

Grell, G. A., J. Dudhia, and D. R. Stauffer, 1995: A description of the fifth-generation Penn State/NCAR Mesoscale Model (MMS). NCAR/TN-398+STR, National Center for Atmospheric Research, Boulder, CO, 138 pp.

Gustafsson, B., H.-O. Kreiss, and J. Oliger, 1995: *Time Dependent Problems and Difference Methods.* Wiley Interscience, 642 pp.

Kreiss, H.-O., 1980: Problems with different time scales for partial differential equations. *Comm. Pure Appl. Math.,* **33,** 399–440.

Lu, C., and G. L. Browning, 1998: The impact of observational and model errors on four-dimensional variational data assimilation. *J. Atmos. Sci.,* **55,** 995–1011.

Talagrand, O., and P. Courtier, 1987: Variational assimilation of meteorological observations with the adjoint vorticity equation. I: Theory. *Quart. J. Roy. Meteor. Soc.,* **113,** 1311–1328.

Tsuyuki, T., 1996: Variational data assimilation in the tropics using precipitation data. Part I: Column model. *Meteor. Atmos. Phys.,* **60,** 87–104.

Verlinde, J., and W. R. Cotton, 1993: Fitting microphysical observations of non-steady convective clouds to a numerical model: An application of the adjoint technique of data assimilation to a kinematic model. *Mon. Wea. Rev.,* **121,** 2276–2793.

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

Unauthenticated | Downloaded 08/04/21 12:43 PM UTC

——, 1996b: Generalized adjoint for physical processes with parameterized discontinuities. Part II: Vector formulations and matching conditions. *J. Atmos. Sci.,* **53,** 1143–1155.
——, 1997a: Generalized adjoint for physical processes with parameterized discontinuities. Part III: Multiple threshold conditions. *J. Atmos. Sci.,* **54,** 2713–2721.
——, 1997b: Generalized adjoint for physical processes with parameterized discontinuities. Part IV: Problems in time discretization. *J. Atmos. Sci.,* **54,** 2722–2728.
Zou, X., 1997: Tangent linear and adjoint of ''on-off'' processes and their feasibility for use in 4-dimensional variational data assimilation. *Tellus,* **49A,** 3–31.
——, I. M. Navon, and J. G. Sela, 1993: Variational data assimilation with moist threshold processes using the NMC spectral model. *Tellus,* **45A,** 370–387.
Zupanski, D., 1993: The effect of discontinuities in the Betts–Miller cumulus convection scheme on four-dimensional variational assimilation. *Tellus,* **45A,** 511–524.
——, and F. Mesinger, 1995: Four-dimensional variational assimilation of precipitation data. *Mon. Wea. Rev.,* **123,** 1112–1127.

Unauthenticated | Downloaded 08/04/21 12:43 PM UTC