Acta Oceanol. Sin., 2020, Vol. 39, No. 9, P. 140–154
https://doi.org/10.1007/s13131-020-1654-5
http://www.hyxb.org.cn
E-mail: hyxbe@263.net

A variational successive corrections approach for the sea ice concentration analysis

Xuefeng Zhang[1], Lu Yang[1], Hongli Fu[2]*, Dong Li[2], Zheqi Shen[3], Lianxin Zhang[2], Xuhui Hu[1]

[1] School of Marine Science and Technology, Tianjin University, Tianjin 300110, China

[2] Key Laboratory of Marine Environmental Information Technology, National Marine Data and Information Service, Ministry of Natural Resources, Tianjin 300171, China

[3] Second Institute of Oceanography, Ministry of Natural Resources, Hangzhou 310012, China

Received 3 June 2019; accepted 13 January 2020

Abstract

The sea ice concentration observation from satellite remote sensing includes the spatial multi-scale information. However, traditional data assimilation methods cannot better extract the valuable information due to the complicated variability of the sea ice concentration in the marginal ice zone. A successive corrections analysis using variational optimization method, called spatial multi-scale recursive filter (SMRF), has been designed in this paper to extract multi-scale information resolved by sea ice observations. It is a combination of successive correction methods (SCM) and minimization algorithms, in which various observational scales, from longer to shorter wavelengths, can be extracted successively. As a variational objective analysis scheme, it gains the advantage over the conventional approaches that analyze all scales resolved by observations at one time, and also, the specification of parameters is more convenient. Results of single-observation experiment demonstrate that the SMRF scheme possesses a good ability in propagating observational signals. Further, it shows a superior performance in extracting multi-scale information in a two-dimensional sea ice concentration (SIC) experiment with the real observations from Special Sensor Microwave/Imager SIC (SSMI).

Key words: variational successive corrections, spatial multi-scale recursive filter, sea ice concentration

Citation: Zhang Xuefeng, Yang Lu, Fu Hongli, Li Dong, Shen Zheqi, Zhang Lianxin, Hu Xuhui. 2020. A variational successive corrections approach for the sea ice concentration analysis. Acta Oceanologica Sinica, 39(9): 140–154, doi: 10.1007/s13131-020-1654-5

1 Introduction

Objective analysis, a kind of techniques for gridding observations, has developed and evolved for many decades. Historically, its main purpose is to provide initial conditions for operational prediction models and aid in the diagnostic studies in atmospheric and oceanic field. Although the use for the first purpose has all but disappeared today due to the springing-up of other more sophisticated schemes such as optimal interpolation (OI) and variational methods, objective analysis schemes are still widely used for diagnostic purposes and many studies have employed various incarnations of them to investigate a diverse range of researches.

As a branch of objective analysis, successive correction menthod (SCM) represented by Cressman (1959) and Barnes (1964, 1973) have received more attention than others and remain popular today. Cressman (1959) used a series of scans with decreasing radius of influence to retrieve a broad spectrum of wavelengths from the observations. Its major contribution is introducing the practice of building details from longer to shorter waves. Barnes (1964) argued that such a scheme suffered from the disadvantage of tending to smooth out all small variations in the field. Moreover, an unstable iteration may occur in the Cressman scheme and thus an additional dissipation scheme has to be performed (e.g., Seaman, 1983; Seaman and Hutchinson, 1985; Lu and Browning, 1998). To maximize details resolved by observations, Barnes (1964) proposed a scheme similar to the Cressman method but using a different weighting function (Gaussian-type) with the weight factor (radius of influence) fixed for all passes of scan. This algorithm was then replaced by Barnes (1973) employing only two passes (one initial pass and one correction pass) with a diminished smoothing factor for the second (correction) pass. The most attractive feature of the Barnes scheme is its well-known response characteristics. By choosing the smoothing parameters, one can ascertain which range of wavelengths will be retained in the final analysis. However, the response function of the Barnes (1964, 1973) is derived under the assumption that the observations are continuous and unbounded (infinite). Practically, it is best applicable to reasonably uniform data distributions. If data are irregularly distributed, the phase of the response function will change and signals may be distorted in the analyzed field. The more sparsely and irregularly distributed the data, the less results are in accordance with the theoretical predictions (Achtemeier, 1986; Pauley and Wu, 1990; Buzzi et al., 1991). This raises the difficulty of choosing the appropriate parameters in a particular application to achieve an optimum analysis. In fact, in the situation of sparse and irregular data distribution, no single selection of the parameters can produce the most accurate analysis for all wavelengths.

Foundation item: The National Key Research and Development Program of China under contract Nos 2017YFC1404103 and 2016YFC1401701; the National Programme on Global Change and Air-Sea Interaction of China under contract GASI-IPOVAI-04; the National Natural Science Foundation of China under contract Nos 41876014 and 41606039.
*Corresponding author, E-mail: fhlkjj@163.com

As we know, an effective objective analysis scheme should at least be able to retrieve resolvable long wavelengths in data-spare areas and preserve details in data-dense areas. If multi-wavelengths are extracted simultaneously without an effective mechanism, the analysis can be seriously contaminated by noises, arising from observational errors or irregular data distribution. A practical way to relieve this problem is analyzing first for larger scales and then for shorter scales. The more accurate the long wavelengths, the less impact the noises may have on the analyzed field. Therefore, when applied to irregularly spaced data, it is advisable for analysis scheme sequentially decrease the smoothing factors as Cressman method does in order to retain the most accurate analysis of the longer wavelengths. Based upon this idea, some variational successive corrections approaches, such as the multi-grid approach, the multi-scale diffusion filter approach, etc, have proposed by scholars as listed in Table 1. Among them, Xie et al. (2005, 2010) proposed a multi-scale 3D-VAR implemented by the multi-grid technique, using a sequence of grids with different resolutions to correct different wavelengths. The analysis is interpolated between two consecutive grid level and then enter into a new analysis cycle. Strictly speaking, the multi-grid 3D-VAR is not a traditional sense of 3D-VAR because it solves a series of 3D-VAR respectively on different grid levels. If the background filed is neglected and only the observational field is considered, the multi-grid 3D-VAR then evolves into a variational objective analysis scheme. Then one problem arise, is it possible for a variational objective analysis method to handle all spatial scales of observations in a single iterative procedure, rather than solve certain number of variational problems?

In this paper, a variant of SCM that can satisfy the above requirement, called SMRF, is proposed. Its main idea persists with other successive correction schemes in extracting multi-scale information from observations. Unlike these other schemes, it uses a variational optimization technique to minimize the difference between the estimated and the observed field. It is actually a combination of SCM and a minimization algorithm. We incorporate scales information into a minimization algorithm by using a recursive filter at each iteration to retrieve desired wavelengths successively. As a result, apart from the advantage in multi-scale information extraction, this scheme gains extra benefits from the minimization procedure: first, the inherent convergence property is guaranteed; second, the weighting parameters can be automatically determined by a line search algorithm without manual interventions; the last, it can analyze the data in all scales at one time.

The paper is organized as follows. The background knowledge related to the topic of this paper is briefed in Section 2, including the SCM and gradient-based minimization algorithms. In Section 3, in view of the relationship between the SCM and gradient-based minimization algorithms, the SMRF scheme is proposed by incorporating scales information into a minimization procedure. In Section 4, a single-observation experiment and an idealized SIC assimilation experiment are performed to evaluate the new scheme. The conclusions are summarized in Section 5.

2 Necessary background

In this section, the SCM and the gradient-based minimization algorithms are briefly introduced.

2.1 *Basics of SCM*

The SCM is a kind of empirical approaches to correct the first-guess field by a linear combination of residual difference between the predicted and the observed values. In other words, the initial estimation field is gradually modified with the actual observation field until the correction factor is no greater than the given error value, at which time the revision process can be considered as the end. The formula is as follows:

$$G_{i,j}^{n+1} = G_{i,j}^{n} + C_{i,j}^{n}, \tag{1}$$

where i,j is index of the analyzed grid point, the superscript n denotes the n-th iteration, $G_{i,j}^{n}$is the analyzed value at the n-th iteration, and $G_{i,j}^{0}$ corresponds to the first-guess field, $C_{i,j}^{n}$ is the correction factor. The expression of correction factor is:

$$C_{i,j}^{n} = \sum_{s=1}^{M} W_s^n Q_s \Big/ \sum_{s=1}^{M} W_s^n, \tag{2}$$

where M is the total number of observations within the circular areas of i,j point with radius R, s denotes the s-th observation, Q_s is difference between the predictedand the observed values, and W_s^n is a weight function at the n-th iteration. The SCM can correct an analysis from longer to shorter wavelengths by changing the weight functions during iterations.

2.2 *Basics of gradient-based minimization methods*

The basic problem is to minimize a cost function as follows

$$\min J(x),$$

where x is the controlling variable, typically $x \in R^n$, but this can also be subject to constraints. To numerically approximate the solution, a sequence $\{x_n\}_{n=1}^{\infty}$ should be constructed so that $x_n \to x^*$, where $J(x^*) = \min J(x)$. Many kinds of algorithms existed for this problem, one is known as gradient-based, in which the sequence $\{x_n\}_{n=1}^{\infty}$ is constructed iteratively by choosing a search direction p_n at each iteration and minimizing $J(x)$ along this direction. This reduces the problem essentially to a sequence of

Table 1. Existing variational successive corrections approach

Paper information	Approach	Explanation
He et al., 2008	sequential 3D-VAR method	deal with data globally with an explicit specification of the observation errors
Li et al., 2008	multi-grid data assimilation scheme	quickly minimizing longwave and shortwave errors successively
Peng et al., 2010	scale-selective data assimilation	correct large scale deviations in regional climate models
Xie et al., 2010	sequential variational analysis approach	multi-scale, inhomogeneous, anisotropic, and temporally consistent
Li et al., 2011	multi-scale diffusion filter	extract multi-scale information of ocean observation data
Wu et al., 2011	3D-VAR based on sequential filter	retrieve structures of true field appropriately
Zhang et al., 2014	background adjustment scheme	suitable for irregular interval climate observation data
Zhang et al., 2015	spatial multi-scale diffusion filters	retrieve the longwave information over the whole analysis domain and the shortwave information over data-dense regions

one-dimensional problem and x_n is given by the basic recurrence:

$$x_{n+1}=x_n+l_n p_n, \tag{3}$$

where l_n is the step length, p_n is usually constructed using gradient information and we call p_n a descent direction if $\nabla J_n \cdot p_n<0$, where ∇J_n is the gradient of J with respect to x at x_n. Based on the conjugate gradient optimization theory, p_n is fomulated as a product of ∇J_n and a positive definite matrix E_n, namely $p_n=-E_n\nabla J_n$. If E_n is simplified to a unit matrix, Eq. (3) is the well-known steepest descent algorithm. Once the descent direction p_n is selected, the step length l_n can be determined through a line search algorithm (Moré and Thuente, 1994) to insure a sufficient decrease of the cost function along this direction.

3 The SMRF scheme

To retrieve multi-scale information resolved by observations, a variant of SCM scheme using variational technique, called SMRF, is developed. It is a combination of SCM and minimization algorithms.

3.1 *Similarities between the SCM and the minimization algorithm*

Actually, the recursion formulated by Eq. (3) is also a procedure of successive correction. Considering the following problem that minimizes the difference between the estimated and the observed values

$$\min J(x) = \min \frac{1}{2}(x^{\mathrm{o}} - \boldsymbol{H}x)^{\mathrm{T}}\boldsymbol{R}^{-1}(x^{\mathrm{o}} - \boldsymbol{H}x)\,, \tag{4}$$

where x is the analyzed field, x^{o} is the observed field, $\boldsymbol{H}$ is an interpolation operator from analysis space to observation space, $\boldsymbol{R}$ is the observational error covariance matrix, $(\cdot)^{\mathrm{T}}$ indicates transpose, and $(\cdot)^{-1}$ indicates inversion, the gradient of $J(x)$ is:

$$\nabla J(x) = -\boldsymbol{H}^{\mathrm{T}}\boldsymbol{R}^{-1}(x^{\mathrm{o}} - \boldsymbol{H}x). \tag{5}$$

Apparently, $\nabla J(x)$ actually represents the residual difference between the observed value x^{o} and the estimated value x on analysis grid. In Eq. (3), if we choose $p_n = -\boldsymbol{E}_n\nabla J(x_n)$ ($\boldsymbol{E}_n$ is a positive definite matrix) as the descent direction, Eq. (3) becomes

$$x_{n+1} = x_n + w_n\,(x^{\mathrm{o}} - \boldsymbol{H}x_n)\,, \tag{6}$$

where $w_n = l_n\boldsymbol{E}_n\boldsymbol{H}^{\mathrm{T}}\boldsymbol{R}^{-1}$ is the weight for the n-th iteration. Equation (6) has the same form as the successive correction procedure except that the weights w_n are different and are obtained in different ways.

3.2 *Problems with minimization algorithms*

Once the gradient is obtained according to Eq. (5), the problem Eq. (4) then can be solved by using such a minimization algorithm as the steepest descent, the LBFGS, or the conjugate gradient method. However, this problem is usually ill-posed due to the scarcity and the irregular distribution of the observations. Further, without an effective mechanism of transmitting observational signals, the analysis will lose its coherent long-wave feature in data-void areas. From a minimization viewpoint, we will reveal in this portion that the underlying cause lies in the "flawed" gradient $\nabla J(x)$ arising from the irregular data distribution.

For simplicity, $\boldsymbol{R}$ is assumed to be an identity matrix, Eq. (5) then becomes

$$\nabla J(x) = -\boldsymbol{H}^{\mathrm{T}}\,(x^{\mathrm{o}} - \boldsymbol{H}x)\,. \tag{7}$$

Given n analyzed grid points and m observational locations, then x is a vector of length n, x^{o} is a vector of length m. We also assume that the observations are right located at the analyzed grid points and $m<n$. In such a case, the analyzed grid points can always be indexed by a certain order of observational locations so that $\boldsymbol{H}$ has the following form

$$\boldsymbol{H}_{m\times n} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & \dots & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & \dots & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & 0 & 1 & 0 & \dots & 0 \end{bmatrix}. \tag{8}$$

Note that the last n-m columns of $\boldsymbol{H}$ are all zero vectors. Accordingly, the last n-m elements of $\boldsymbol{H}^{\mathrm{T}}\,(x^{\mathrm{o}}-\boldsymbol{H}x)$ are zero elements. As a result, for a grid point where no measurements are available, the corresponding element of $\boldsymbol{H}^{\mathrm{T}}\,(x^{\mathrm{o}}-\boldsymbol{H}x)$ at that position is definitely equal to zero, while for those observed grid points, the corresponding elements of $\boldsymbol{H}^{\mathrm{T}}\,(x^{\mathrm{o}}-\boldsymbol{H}x)$ remain their actual values. That is, the distribution of $\boldsymbol{H}^{\mathrm{T}}\,(x^{\mathrm{o}}-\boldsymbol{H}x)$, and thus $\nabla J(x)$, is spatially incoherent. Though from a mathematical viewpoint, $\nabla J(x)$ obtained in this way is unquestionable, this phenomenon is unreasonable in a physical sense because it is merely caused by the irregular data distribution.

If this "flawed" gradient is introduced into a general gradient-based minimization algorithm, it's not strange that the analysis will deviate far from what we anticipate. Taking the steepest descent algorithm, for example, the estimate is updated at the i-th iteration by

$$x_i = x_{i-1} + l\,(-\nabla J(x_{i-1}))\,, \tag{9}$$

where l is the step length obtained by a line search algorithm. Supposing $x=x_0$ is the initial guess, the first iteration yields

$$x_1 = x_0 + l\,(-\nabla J(x_0))\,.$$

As indicated above, $\nabla J(x_0)$ is spatially incoherent in data-void areas and therefore x_1 will also involve amounts of erroneous small scales in these regions. The same issue runs through all later iterations, leading to a long-wave loss in data-void areas. As for those other gradient-based minimization algorithms, such as the LBFGS and the conjugate gradient method, the same problem exists, for a similar reason.

3.3 *Variational form of SCM*

Recognizing the defect of the conventional minimization algorithms in solving an ill-posed problem, and recalling the resemblance of a SCM and a minimization algorithm, we are enlightened to refer to the desirable feature of a SCM scheme in multi-scale analyzing and incorporate it into a minimization algorithm. We apply a recursive filter to the gradient of the cost function at each iteration of a minimization procedure. With the filter parameter decreasing sequentially with iterations, various scales, from longer to shorter wavelengths, can be extracted successively (see Appendix for recursion details).

We now give a brief analysis on the fundamentals of this

scheme. The gradient $\nabla J(x)$ described by Eq. (5) actually represents observational residuals at x. The scheme starts by applying a recursive filter $\boldsymbol{E}$ to $-\nabla J(x_0)$ with a large enough α, resultant $\boldsymbol{E}(-\nabla J(x_0))$ then reasonably characterizes the "longest" wavelengths of the observational residuals at x_0. Also, since the recursive filtering operator $\boldsymbol{E}$ is positive definite, $\boldsymbol{E}(-\nabla J(x_0))$ is guaranteed to be a descent direction which insures the decrease of the residual difference between the estimated and the observed values along this direction. However, just as what we have depicted in the first part of Appendix, for any wavelength the filtering process of $\boldsymbol{E}(-\nabla J(x_0))$ will lead to some amplitude loss. A reasonable analysis over data-sparse areas requires the long waves be captured as accurately as possible so that it will not interfere with the extraction of shorter wavelengths in later iterations. Therefore, to regain some of those lost information, a line search procedure is performed along this direction to find an appropriate step length l. When the estimate is updated by $x_1 = x_0 + l\boldsymbol{E}(-\nabla J(x_0))$, the "largest" scale of the observational residuals at $x = x_0$ is "fully" extracted and incorporated into the new estimate x_1. Then α is diminished appropriately, as a result, the "largest" scale of the observational residuals at $x = x_1$ can be captured at the second iteration and incorporated into x_2. As iteration proceeds, all scales, from longer to shorter wavelengths, can be pulled out successively.

Actually, this scheme is a natural extension of the Barnes SCM scheme (Barnes, 1964, 1973). But it is in a variational form with the advantage that the weights can be automatically obtained by a line search algorithm. This scheme can also be regarded as a minimization algorithm which gains an advantage over conventional minimization algorithms by accounting for various spatial scales resolved by the observations.

To further suppress observational noises, we make a slight modification to our scheme by replacing the problem described by Eq. (4) with the following:

$$\begin{cases} \min J(w) = \min \dfrac{1}{2}(x^o - \boldsymbol{H}\boldsymbol{B}w)^{\mathrm{T}} R^{-1}(x^o - \boldsymbol{H}\boldsymbol{B}w) \\ x = \boldsymbol{B}w \end{cases}, \qquad (10)$$

where $\boldsymbol{B}$ is another recursive filtering operator with a very small filter parameter β. Obviously, problem Eq. (4) is the special case of Eq. (10) when $\beta = 0$. For the same reason as we have explained, solving Eq. (10) directly using a conventional minimization algorithm (e.g., the steepest descent, the LBFGS and the conjugate gradient method) may not yield a well-behaved analysis, as is verified by our experiments in Section 4. Our algorithm is modified as a flow chart shown in Fig. 1.

It should be noted that the cost function defined by Eq. (10) is a counter part of that used in a 3D-VAR, representing the observational term. Therefore, the way we used in analyzing for the gradient is also suitable for a 3D-VAR scheme, see Appendix for details.

4 Experiment designs and results

4.1 *Single-observation experiment*

An effective mechanism for transmitting observational information should be able to: (1) insure the accuracy of the analysis. (2) make observational signals propagate to more wide areas so that the analysis for long waves can be dramatically improved in data-sparse or data-void regions. To test the ability of the SMRF scheme, two experiments are carried out using a single observation. In the first experiment, the recursive filtering operator $\boldsymbol{B}$ with an invariant filter parameter β is applied to the w and Eq. (10) is directly solved using a conventional LBFGS minimization algorithm (Liu and Nocedal, 1989). In the second experiment, another recursive filtering operator $\boldsymbol{E}$ with variable filter parameters is applied to the cost function gradient $-\nabla J(w)$ besides the recursive filtering operator $\boldsymbol{B}$ is applied to w, and Eq. (10) is solved using the SMRF scheme, as shown in Fig. 1.

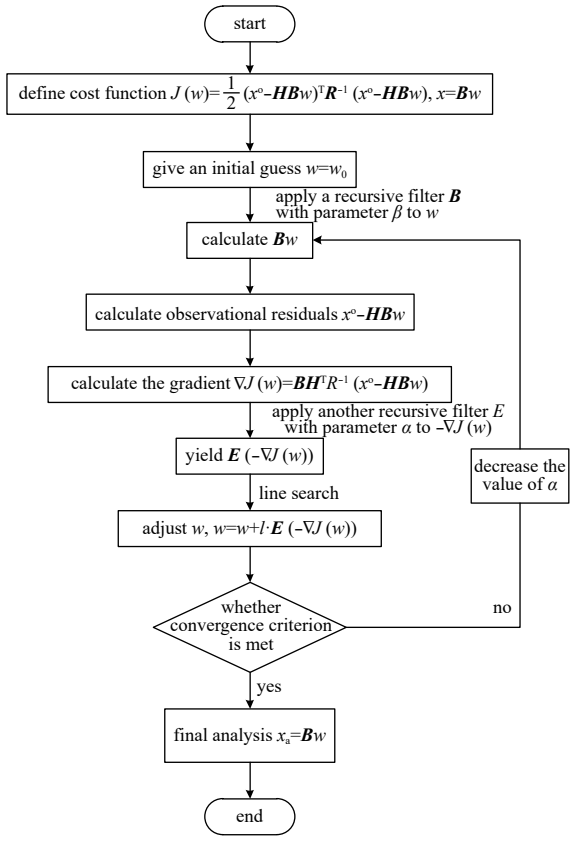

Fig. 1. A flow chart of SMRF process.

4.1.1 *Data and parameters*

The analysis domain covers a square region, extending 10° both in latitude and in longitude. The grid resolution is 0.25°×0.25°. We place only one observation with its value equal to 1.0 at the center of this domain. The number of filtering passes M is set to 8. The filer parameter α in our scheme is chosen as the following Gaussian function:

$$\alpha = \alpha_{\max} \cdot \mathrm{e}^{-\frac{i^2}{2\sigma^2}}, \qquad i = 0, 1, \cdots, N, \qquad (11)$$

where i represents the iteration number, N is a constant to be set, $\sigma = \dfrac{N}{4}$, $\alpha_{\max} = 0.999$. As we can see, at the beginning, $\alpha = \alpha_{\max}$ $(i = 0)$, then α decreases with iterations and almost approaches zero (exactly, e^{-8}) when $i = N$, so the number of iterations can be chosen to be no greater than N in practical implementations. In this experiment we set $N = 250$. The observation operator $\boldsymbol{H}$ is a simple bilinear interpolation. The initial-guess field w_0 is selected to be zero. The line search algorithm is based on the study by Moré and Thuente (1994).

4.1.2 *Results*

Figure 2 shows the results of solving Eq. (10) by the LBFGS al-

gorithm when filter parameter β=0.1 and 0.4, respectively. As can be seen, since a recursive filter makes grid points connect and interact with each other, even a single observation can transmit observational signals to neighboring grid points. Different choice of β will yield different analysis. If β is relatively large, the observational signals can propagate to more wide range of areas, but the analysis will lose accuracy in practical use. If β is small (e.g., β=0.1), the analysis approaches the observation closely but remains all most unchanged in data-void regions because observational signals cannot propagate there. Thus, maximizing the details requires a small β and filling in the data-void areas with long waves needs a large β. It seems that there is no way to take care of both of these two aspects simultaneously. However, this can to some extent be remedied by our scheme.

Figure 3 tells the results of the SMRF scheme at different iterations using a small filter parameter β(=0.1). Figures 3a, b, c and d are for iteration 50, 100, 130, and 180, respectively. Apparently, the analysis is corrected from large scales to details. As a result,

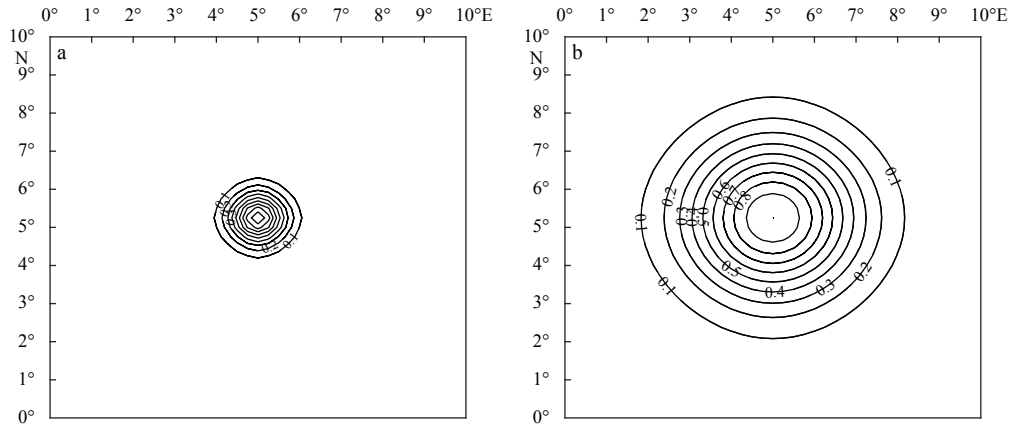

Fig. 2. The spread of observational information using the LBFGS algorithm when β=0.1 (a) and β=0.4 (b).

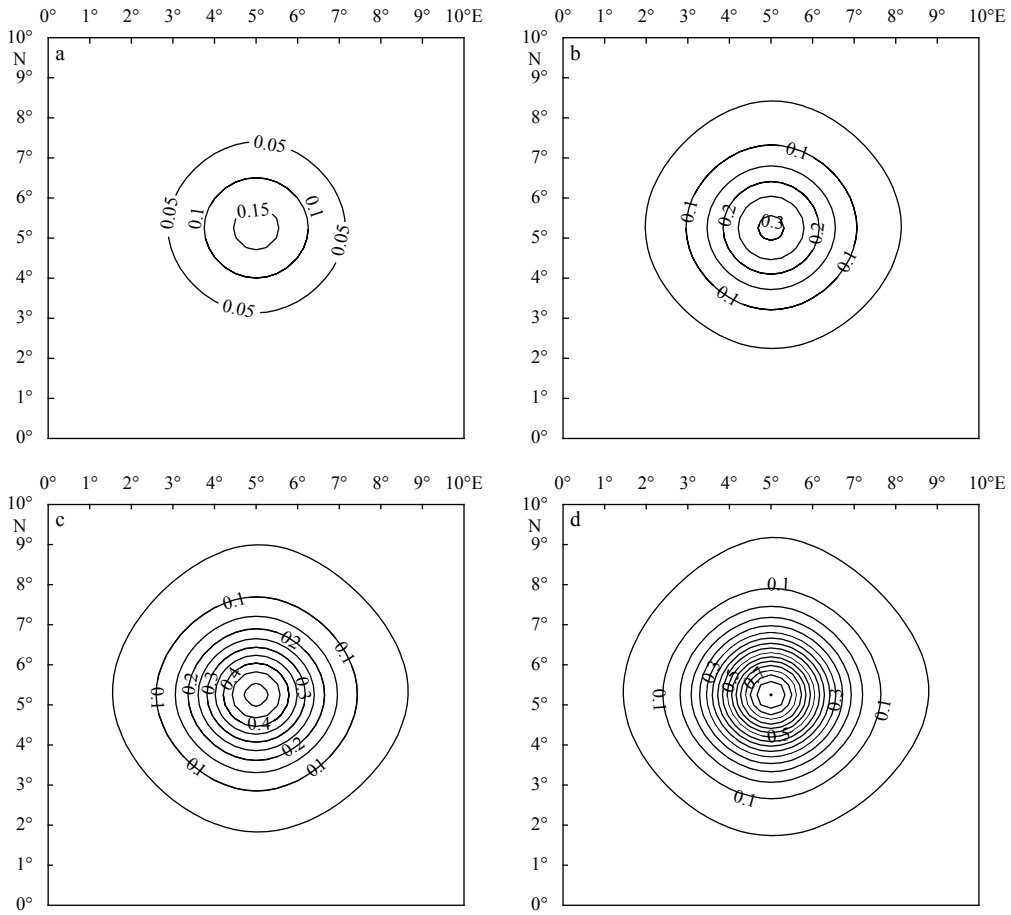

Fig. 3. The spread of the observational information in the SMRF scheme when β=0.1. a, b, c and d. The results at iteration 50, 100, 130 and 180, respectively.

observational signals can propagate to more wide regions while at the same time the analysis does not lose its accuracy. For a further understanding of the evolving process of our analysis with iterations, the surface plots of Figs 2a and 3 are presented as Figs 4 and 5 respectively. Figure 4 shows that using the LBFGS algorithm with a small β(=0.1), while the accuracy is guaranteed near the observed locations, the observation can only have an effect on a very close area around it. Figure 5 depicts that in the SMRF scheme, with the same small β(=0.1), the analysis starts with a coarse field and approaches the observed value gradually, and the observational signals can be transmitted to more wide areas compared with that in Fig. 4.

4.2 *Application to SIC experiment*

For further verifying the effectiveness of the SMRF scheme in extracting spatial multi-scale information, a two-dimensional experiment with SSMI SIC observations is carried out. To reveal how different wavelengths are corrected sequentially, the analyzed field and the descent direction at different iterations in the SMRF scheme are explored in comparison with the counterparts solved by using the steepest descent algorithm.

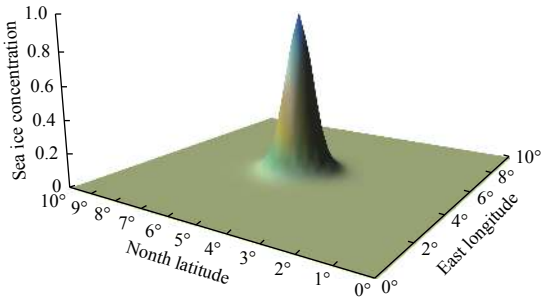

Fig. 4. The surface plot of Fig. 2a.

4.2.1 *Data and parameters*

The SSMI daily SIC data are obtained from the National Snow and Ice Data Center (NSIDC), the horizontal resolution of which is 25 km × 25 km. The analysis domain covers the Arctic Ocean. The "true" state of SIC field is shown in Fig. 6a, which is constructed by the SIC observations from the SSMI on September 1, 2014. Since the spatial resolution of the analysis field is usually different from the satellite observation, we select one observation for every four analysis grid points. We also remove partial points located in the sea ice marginal ice zone to examine the validation of the SMRF scheme. As a consequence, there are 1 384 observations (Fig. 6b) remained to restore the "true" field. The observation errors are assumed to be uncorrelated and therefore a diagonal matrix is used with all diagonal elements equal to the square of the observation standard deviation, σ_o^2. σ_o here has been normalized to 1.0 in order to avoid the complexity. β is chosen to be 0.2. N is set to 500. The other settings remain the same as those in the single-observation experiment above.

4.2.2 *Results of the steepest descent algorithm*

As can be seen from Fig. 7, the analysis results constructed by the steepest descent algorithm deeply rely on β. The small (large) β, which is related to the small (large) radius of influence, only reflects the short (long) wave information of the observations, indicating the long and short wave information cannot be resolved

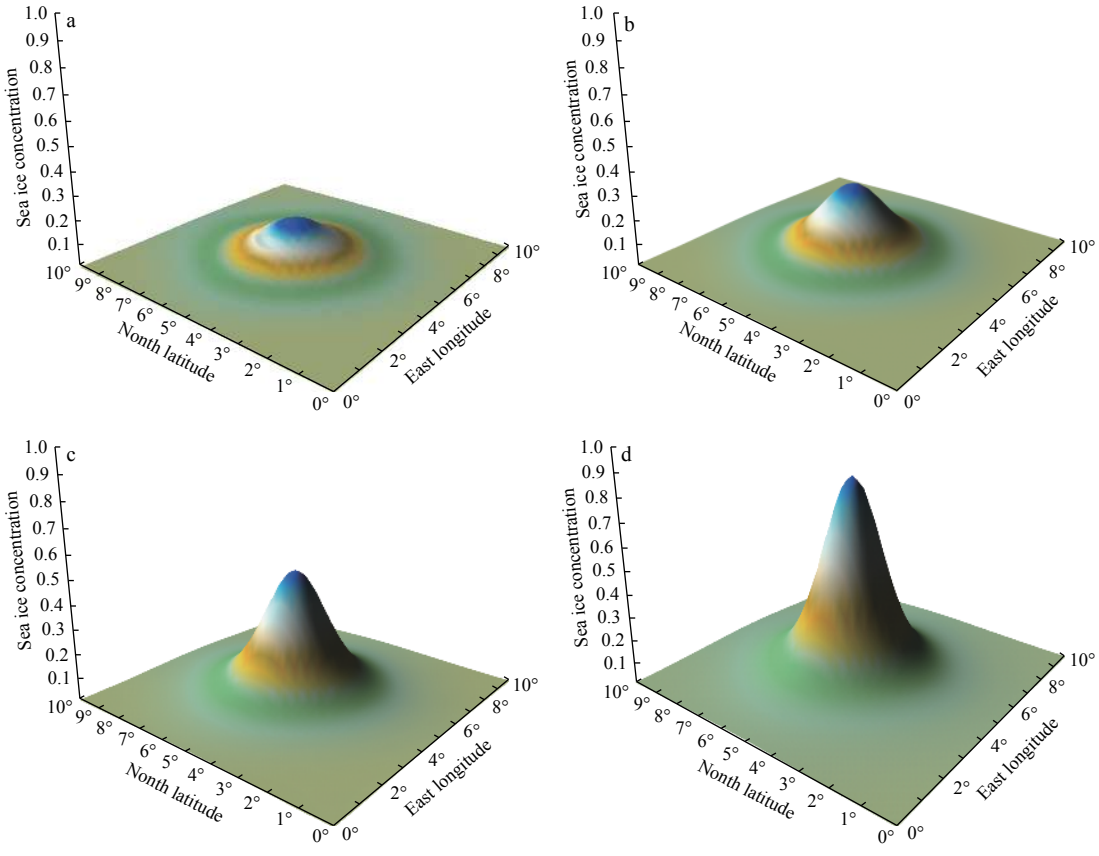

Fig. 5. The surface plot of Fig. 3.

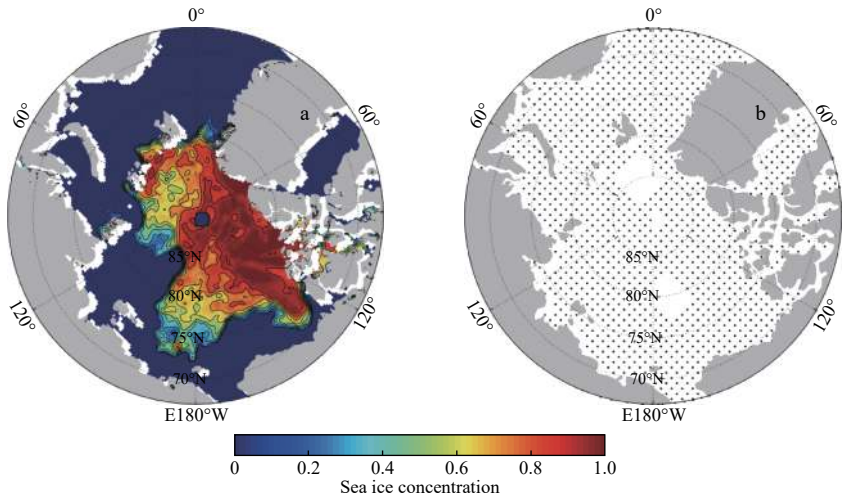

Fig. 6. The true SIC field of Arctic Ocean constructed based on the SSMI SIC on September 1, 2014 (a); and the locations of "observations" (b).

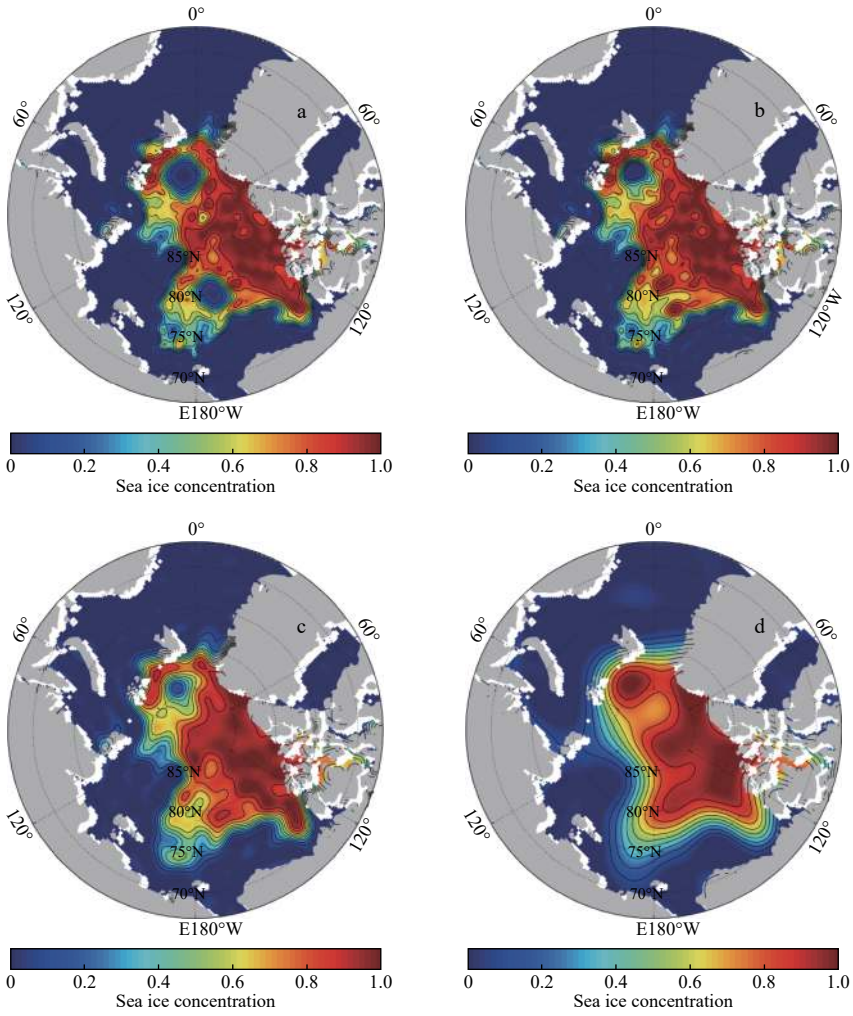

Fig. 7. Analyzed field solved by using the steepest descent algorithm β=0.2 (a), β=0.4 (b), β=0.6 (c) and β=0.8 (d), in which the iteration is 25, 162, 165 and 148, respectively.

simultaneously. Figure 8 shows results of the steepest descent algorithm with β=0.2 at iteration 3, 5 and 7. The descent direction is spatially incoherent in the data-void region because of the sharp variation of the gradient caused by the irregular distribution of observations (Figs 8b, d and f). Accordingly, there analysis updated along this direction tend to the incoherent structure (Figs 8a,

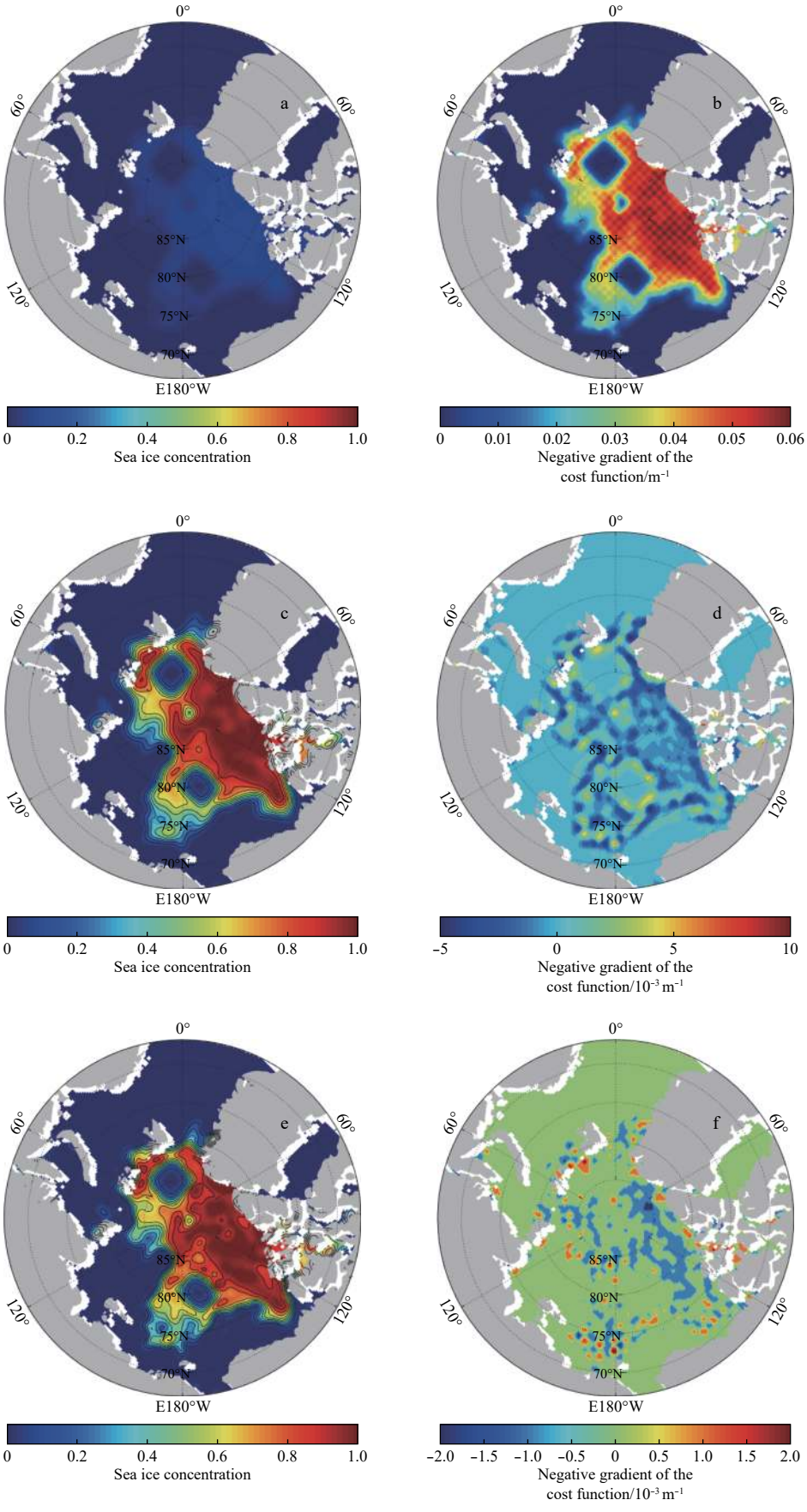

Fig. 8. Analyzed field (left column) and the descent direction ($-\nabla J$) (right column) solved using the steepest descent algorithm (β=0.2) at iteration 3, 5 and 7, respectively.

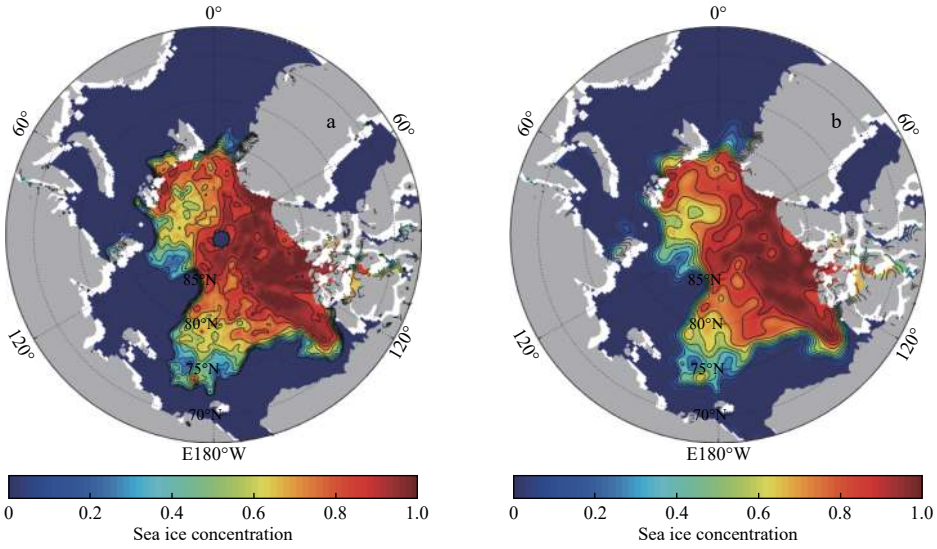

Fig. 9. The true SIC (a) and the analysis result (b) from the SMRF scheme with β=0.2 and N=500.

c and e). The same problem will exist for other gradient-based minimization algorithms such as the quasi-Newton methods, LBFGS and the conjugate gradient method.

4.2.3 *Results of the SMRF scheme*

The analysis results from SMRF scheme with β=0.2 is very similar with the true field (Fig. 9), which avoids the incoherent spatial structure in the data-void area compared to the steepest descent algorithm. Therefore, the SMRF scheme can better account for various spatial scales resolved by the observations, and the long and short wavelength information can be extracted simultaneously from the observations (Figs 10a, c and e). This is attributed to the fact that the descent direction is built by smoothing out the sharp variation of the gradient to extract the long wave of the observational residuals. As the filtering scales α decreases with iterations, the descent direction is obtained from longer to shorter wavelengths (Figs 10b, d and f). Consequently, the analyzed field adjusted along this direction can also be extracted successively.

4.2.4 *Choice of N*

Ideally, α should decrease continuously with iterations. However, discrete ones are needed in practical implementations. As α takes the form of Eq. (11) in our scheme, the choice of N is an issue to be considered. Figure 11 gives the analysis results with different value of N. As can be seen, the desirable analysis can be achieved as $N>150$ in this experiment. It is also shown in our other experiments that the choice of N is not a problem in practice because we can usually achieve reasonable results as long as N is big enough. However, too big value is not necessary and also not recommended because of the computational cost.

5 Conclusions and discussion

In this study, a muti-scale variational optimization technique is designed to extract spatial multi-scale information resolved by observations. In view of the similarity in form between the SCM schemes and the gradient-based algorithms, the new approach incorporates scales into the minimization algorithms. Additionally, to propagate observational signals, it applies recursive filters to the gradient of the cost function and makes filtering scales decrease with iterations to extract various scales. Based on SRMF scheme, the SIC analysis fields can be successfully reconstructed through extracting the information of the real SSMI from long to short waves in turn.

The main conclusions can be summarized as follows.

(1) This scheme is a variant of conventional SCM that can better account for resolvable multi-scale in the observations. Actually, it is a natural extension of the Barnes scheme but in a variational form, which brings us several extra benefits. First, the specification of scheme parameters is relatively easy because the weights are automatically determined by a line search algorithm. Second, the convergence is implied in a minimization procedure and the "distance" between the estimate and the observed value is diminished with iterations. The last, all wavelengths are analyzed at one time in a single iterative procedure.

(2) From a physical viewpoint, the spatial distribution of the gradient of a cost function defined in a variational problem may be unreasonable, for example, in condition that data are irregularly distributed. Use of this gradient in a conventional minimization algorithm (e.g., the steepest descent, the LBFGS, or the conjugate gradient method) will cause a poor analysis. Our scheme is a remedial approach for this issue. Though inter-comparison studies are performed in our experiments between the conventional minimization algorithms and the SMRF scheme, this is not our real purpose because it is unfair for these algorithms in solving an ill-conditioned problem. On the contrary, we simply intend to present the feasibility and the effectiveness of the combination of SCM and a minimization procedure in data-spare cases.

(3) Since the cost function defined by Eq. (10) is a counter part of that used in a 3D-VAR, representing the observational term, the problem mentioned in (2) also exists in a 3D-VAR scheme if the background error covariance matrix is not appropriately modeled, which shows the potentiality of the SMRF scheme to be extended to a 3D-VAR, as will be detailed in Appendix.

(4) SMRF aims to capture longer wavelengths as accurate as possible before analyzing for shorter wavelengths. While this can to some extent reduce the chance of the long waves being contaminated by noises, there is no way to completely avoid this.

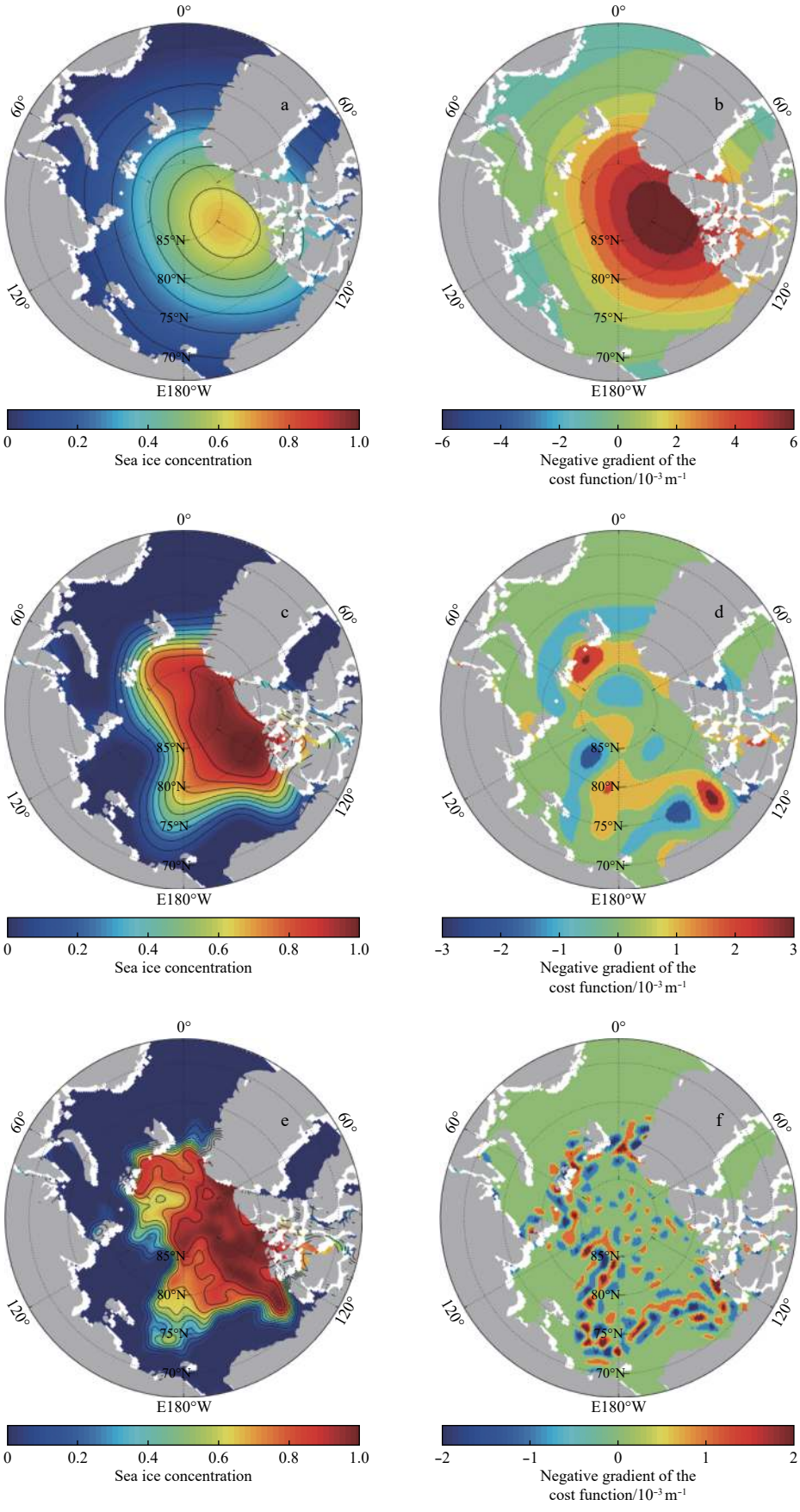

Fig. 10. Analyzed field (left column) and the descent direction (right column) of the SMRF scheme (β=0.2) at iteration 10, 100 and 250.

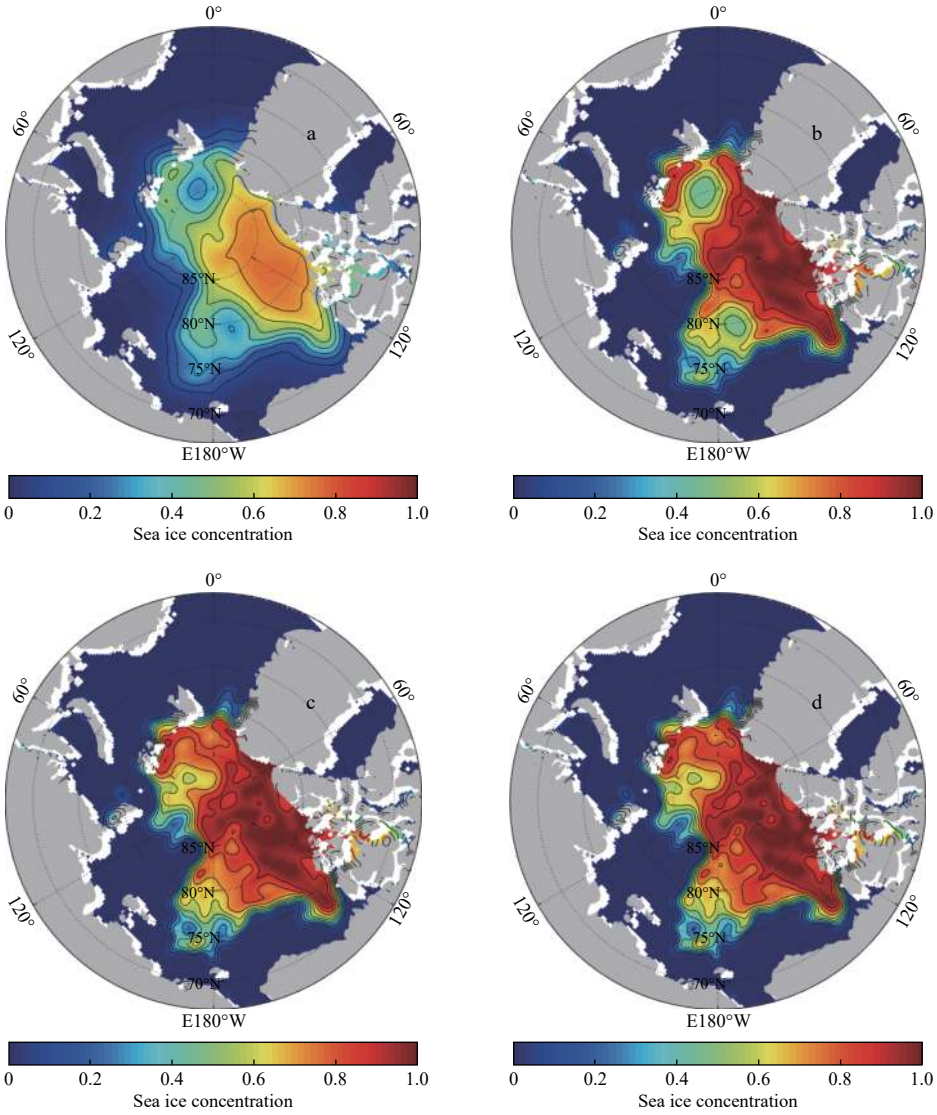

Fig. 11. Analysis of the SMRF scheme (β=0.2) with different choice of N. a. b. c and d. N=10, 20, 150 and 300, respectively.

How much noise contamination is included in an analysis and how much the signals are distorted is still a question to be studied, especially with a quantitative analysis. Additionally, compared with the multi-grid method (Xie et al., 2010), the computational cost is a defect of SMRF, and further improvements are needed.

(5) The high-order recursive filter can effectively avoid the problems such as large truncation error and difficult boundary estimation caused by the cascade of first-order recursive filter used in our study. The high-order recursive filter algorithm will be enclosed in the SMRF in the future. Besides, It will be further investigated to what degree the SMRF can improve the sea ice weather forecast precision and climate prediction skill.

References

Achtemeier G L. 1986. The impact of data boundaries upon a successive corrections objective analysis of limited-area datasets. Monthly Weather Review, 114(1): 40–49, doi: 10.1175/1520-0493(1986)114<0040:TIODBU>2.0.CO;2

Barnes S L. 1964. A technique for maximizing details in numerical weather map analysis. Journal of Applied Meteorology, 3(4): 396–409, doi: 10.1175/1520-0450(1964)003<0396:ATFMDI>2.0.CO;2

Barnes S L. 1973. Mesoscale objective map analysis using weighted time-series observations. Norman: National Severe Storms Laboratory, 60

Behringer D W, Ji Ming, Leetmaa A. 1998. An improved coupled model for ENSO prediction and implications for ocean initialization. Part I: the ocean data assimilation system. Monthly Weather Review, 126(4): 1013–1021, doi: 10.1175/1520-0493(1998)126<1013:AICMFE>2.0.CO;2

Buzzi A, Gomis D, Pedder M A, et al. 1991. A method to reduce the adverse impact that inhomogeneous station distributions have on spatial interpolation. Monthly Weather Review, 119(10): 2465–2491, doi: 10.1175/1520-0493(1991)119<2465:AMTRTA>2.0.CO;2

Courtier P. 1997. Variational methods (gtSpecial IssueltData assimilation in meteology and oceanography: theory and practice). Journal of the Meteorological Society of Japan, 75(1B): 211–218, doi: 10.2151/jmsj1965.75.1B_211

Cressman G P. 1959. An operational objective analysis system. Monthly Weather Review, 87(10): 367–374, doi: 10.1175/1520-0493(1959)087<0367:AOOAS>2.0.CO;2

Derber J, Rosati A. 1989. A global oceanic data assimilation system. Journal of Physical Oceanography, 19(9): 1333–1347, doi: 10.1175/1520-0485(1989)019<1333:AGODAS>2.0.CO;2

Hayden C M, Purser R J. 1995. Recursive filter objective analysis of meteorological fields: applications to NESDIS operational processing. Journal of Applied Meteorology, 34(1): 3–15, doi: 10.1175/1520-0450-34.1.3

He Zhongjie, Xie Yuanfu, Li Wei, et al. 2008. Application of the sequential three-dimensional variational method to assimilating SST in a global ocean model. Journal of Atmospheric and Oceanic Technology, 25(6): 1018–1033, doi: 10.1175/2007JTECHO540.1

Huang Bohua, Kinter J L, Schopf P S. 2002. Ocean data assimilation using intermittent analyses and continuous model error correction. Advances in Atmospheric Sciences, 19(6): 965–992, doi: 10.1007/s00376-002-0059-z

Koch S E, desJardins M, Kocin P J. 1983. An interactive Barnes objective map analysis scheme for use with satellite and conventional data. Journal of Climate and Applied Meteorology, 22(9): 1487–1503, doi: 10.1175/1520-0450(1983)022<1487:AIBOMA>2.0.CO;2

Li Dong, Wang Xidong, Zhang Xuefeng, et al. 2011. Multi-scale 3D-VAR based on diffusion filter. Marine Science Bulletin (in Chinese), 30(2): 164–171

Li Wei, Xie Yuanfu, He Zhongjie, et al. 2008. Application of the multigrid data assimilation scheme to the China Seas' temperature forecast. Journal of Atmospheric and Oceanic Technology, 25(11): 2106–2116, doi: 10.1175/2008JTECHO510.1

Liu D C, Nocedal J. 1989. On the limited memory BFGS method for large scale optimization. Mathematical Programming, 45(1): 503–528, doi: 10.1007/BF01589116

Lorenc A C. 1986. Analysis methods for numerical weather prediction. Quarterly Journal of the Royal Meteorological Society, 112(474): 1177–1194, doi: 10.1002/qj.49711247414

Lorenc A C. 1988. Optimal nonlinear objective analysis. Quarterly Journal of the Royal Meteorological Society, 114(479): 205–240, doi: 10.1002/qj.49711447911

Lorenc A. 1992. Iterative analysis using covariance functions and filters. Quarterly Journal of the Royal Meteorological Society, 118(505): 569–591, doi: 10.1002/qj.49711850509

Lu Chungu, Browning G L. 1998. The impact of observational errors on objective analyses. Journal of the Atmospheric Sciences, 55(10): 1791–1807, doi: 10.1175/1520-0469(1998)055<1791:TIOOEO>2.0.CO;2

Masina S, Pinardi N, Navarra A. 2001. A global ocean temperature and altimeter data assimilation system for studies of climate variability. Climate Dynamics, 17(9): 687–700, doi: 10.1007/s003820000142

Moré J J, Thuente D J. 1994. Line search algorithms with guaranteed sufficient decrease. ACM Transactions on Mathematical Software, 20(3): 286–307, doi: 10.1145/192115.192132

Pauley P M, Wu Xiaihua. 1990. The theoretical, discrete, and actual response of the Barnes objective analysis scheme for one- and two-dimensional fields. Monthly Weather Review, 118(5): 1145–1164, doi: 10.1175/1520-0493(1990)118<1145:TTDAAR>2.0.CO;2

Peng Shiqiu, Xie Lian, Liu Bin, et al. 2010. Application of scale-selective data assimilation to regional climate modeling and prediction. Monthly Weather Review, 138(4): 1307–1318, doi: 10.1175/2009MWR2974.1

Purser R J, Wu Wanshu, Parrish D F, et al. 2003a. Numerical aspects of the application of recursive filters to variational statistical analysis. Part I: spatially homogeneous and isotropic Gaussian covariances. Monthly Weather Review, 131(8): 1524–1535, doi: 10.1175//1520-0493(2003)131<1524:NAOTAO>2.0.CO;2

Purser R J, Wu Wanshu, Parrish D F, et al. 2003b. Numerical aspects of the application of recursive filters to variational statistical analysis. Part Ⅱ: spatially inhomogeneous and anisotropic general covariances. Monthly Weather Review, 131(8): 1536–1548, doi: 10.1175//2543.1

Seaman R S. 1983. Objective analysis accuracies of statistical interpolation and successive correction schemes. Australian Meteorological Magazine, 31(4): 225–240

Seaman R S, Hutchinson M F. 1985. Comparative real data tests of some objective analysis methods by withholding observations. Australian Meteorological Magazine, 33(1): 37–46

Wu Xinrong, Han Guijun, Li Dong, et al. 2011. A three-dimensional variational analysis using sequential filter. In: Proceedings of the 2011 4th International Joint Conference on Computational Sciences and Optimization. Yunnan: IEEE, 1016–1020, doi: 10.1109/CSO.2011.60

Xie Yuanfu, Koch S E, McGinley J A, et al. 2005. A sequential variational analysis approach for mesoscale data assimilation. In: Proceedings of the 21st Conference on Weather Analysis and Forecasting/17th Conference on Numerical Weather Prediction. Washington, DC: American Meteorological Society

Xie Y, Koch S, McGinley J, et al. 2010. A space-time multiscale analysis system: a sequential variational analysis approach. Monthly Weather Review, 139(4): 1224–1240, doi: 10.1175/2010MWR3338.1

Zhang Xuefeng, Li Dong, Chu P C, et al. 2015. Diffusion filters for variational data assimilation of sea surface temperature in an intermediate climate model. Advances in Meteorology, 2015: 751404, doi: 10.1155/2015/751404

Zhang S, Zhao M, Lin S J, et al. 2014. Retrieval of tropical cyclone statistics with a high-resolution coupled model and data. Geophysical Research Letters, 41(2): 652–660, doi: 10.1002/2013GL058879

Appendix:

Spatial recursive filter and spectral analysis

Spatial recursive filters have been widely used in many 3D-VAR schemes to model a background error covariance matrix. A simple homogeneous recursive filter is described as follows (Lorenc, 1992; Hayden and Purser, 1995)

$$\begin{cases} U_i = \alpha U_{i-1} + (1-\alpha) X_i & \text{for } i = 1,2,\cdots,I \\ Y_i = \alpha Y_{i+1} + (1-\alpha) U_i & \text{for } i = I,\cdots,2,1 \end{cases}, \tag{A1}$$

where X_i is the "input" value at the grid point with index i, U_i is the "intermediate" value that is "forward" smoothed with i increasing sequentially, and the "output" value Y_i is smoothed reversely. α is a smoothing parameter that controls the intrinsic scale of the filter ($0<\alpha<1$). Multi-pass filters can be built up by repeated application of Eq. (A1). Previous study (Hayden and Purser, 1995; Purser et al., 2003a) showed that such a filter, when applied with several passes, can approximately model a Gaussian-type filter. The filtering operator constructed in this way is positive definite. A multidimensional filter can be constructed by applying this one-dimensional filter in each direction.

Following the studies by Xie et al. (2005), we now give an analysis on the spectral response of the recursive filter defined by Eq. (A1). A z-transformation of Eq. (A1) yields

$$\begin{cases} \left(1-\alpha z^{-1}\right) \overline{U} = (1-\alpha) \overline{X} \\ (1-\alpha z) \overline{Y} = (1-\alpha) \overline{U} \end{cases},$$

or

$$\frac{\overline{Y}}{\overline{X}} = \frac{(1-\alpha)^2}{1-\alpha(z+z^{-1})+\alpha^2}, \tag{A2}$$

thus, the spectral response of a single pass of filtering as described by Eq. (A1) is

$$D(\alpha,\lambda) = \frac{\overline{Y}}{\overline{X}}\Big|_{z=e^{j\frac{2\pi\Delta x}{\lambda}}} = \frac{(1-\alpha)^2}{1-2\alpha\cos\left(\frac{2\pi}{\lambda}\Delta x\right)+\alpha^2}, \tag{A3}$$

where j represents the imaginary unit and Δx is the grid spacing. When applied with M passes, the spectral response of the recursive filter then becomes

$$D(\alpha,\lambda) = \left(\frac{(1-\alpha)^2}{1-2\alpha\cos\left(\frac{2\pi}{\lambda}\Delta x\right)+\alpha^2}\right)^M. \tag{A4}$$

Usually, bounds on the ratio between the grid spacing Δx and the average data spacing ΔL^* should lie in the range of 0.3–0.5 (Koch et al., 1983). Assuming $\Delta x/\Delta L^* = \frac{1}{2}$ here, then Eq. (A3) is rewritten as

$$D(\alpha,\lambda^*) = \left(\frac{(1-\alpha)^2}{1-2\alpha\cos\left(\frac{\pi}{2\lambda^*}\right)+\alpha^2}\right)^M, \tag{A5}$$

where $\lambda^* = \lambda/2\Delta L^*$, is dimensionless wavelength.

Figure A1 shows the response functions of the recursive filter with M=8 and α=0.2, 0.3, 0.5, 0.7, 0.8 and 0.9, respectively. As can be seen, the characteristics of this filter are very similar to those of a Barnes filter. That is not strange because they are both Gaussian-type. Several points should be noted according to Fig. A1. (1) it is a lower-pass filter tending to filter out all wavelengths under a certain scale. The smoothing parameter α controls the intrinsic scale of the filter. The larger α is, the lower frequency signals are suppressed. (2) Filtering process will cause some loss of amplitude. For any given wavelength, the larger α is, the more magnitude of its amplitude will lose. (3) If α keeps fixed, the shorter wavelengths suffer from a more loss of that amplitude than the longer ones. In order to relieve the amplitude loss in a recursive filter, we use a line search algorithm to regain some of those lost.

Analyzing for the gradient in a 3D-VAR scheme

As we know, a conventional 3D-VAR may not produce a good analysis when data are regularly and sparsely distributed (Xie et al., 2005). Following the similar analysis procedure we have performed, we think the problem of the gradient of a cost function also exists in a 3D-VAR, if the background error covariance matrix are not properly modeled. The following discussion is based upon a recursive

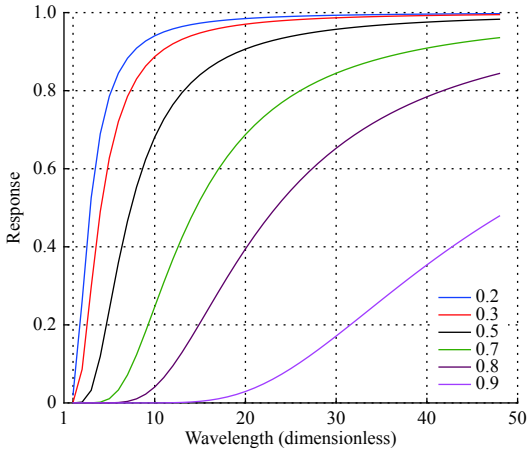

Fig. A1. Response functions of the recursive filter with different value of α.

filtering 3D-VAR.

A general 3D-VAR can be defined as (Courtier, 1997; Lorenc, 1986)

$$\min J(x) = \min \frac{1}{2}(x - x_{\mathrm{b}})^{\mathrm{T}} \boldsymbol{B}^{-1}(x - x_{\mathrm{b}}) + \frac{1}{2}(\boldsymbol{H}x - x^{\mathrm{o}})^{\mathrm{T}} R^{-1}(\boldsymbol{H}x - x^{\mathrm{o}}), \tag{A6}$$

where x_{b} is the background vector, $\boldsymbol{B}$ is the background error covariance matrix, other terms remains the same meaning as represented in Eq. (4) in Section 3.1.

In practical implementation, $\boldsymbol{B}$ can be explicitly generated in a statistical way using correlation scales (Behringer et al., 1998; Derber and Rosati, 1989; Huang et al., 2002; Masina et al., 2001). However, difficulties exist in handling the storage and computation related to $\boldsymbol{B}$ due to its numerous size. To avoid the inversion of $\boldsymbol{B}$ and to speed up the convergence of minimization procedure, a new vector $\boldsymbol{w}$ is introduced by (Lorenc, 1988; Derber and Rosati, 1989)

$$w = \boldsymbol{B}^{-1}(x - x_{\mathrm{b}}).$$

Since $\boldsymbol{B}^{\mathrm{T}}$=$\boldsymbol{B}$, $J(x)$ can be rewritten as

$$J(w) = \frac{1}{2} w^{\mathrm{T}} \boldsymbol{B} w + \frac{1}{2}(\boldsymbol{H}\boldsymbol{B}w - d)^{\mathrm{T}} R^{-1}(\boldsymbol{H}\boldsymbol{B}w - d), \tag{A7}$$

where $d = x^{\mathrm{o}} - \boldsymbol{H}x_{\mathrm{b}}$ is the "innovation" vector. The gradient of $J(w)$ is

$$\nabla J(w) = \boldsymbol{B}\left[w - \boldsymbol{H}^{\mathrm{T}} \boldsymbol{R}^{-1}(d - \boldsymbol{H}\boldsymbol{B}w)\right] = \nabla J_{\mathrm{b}}(w) + \nabla J_{\mathrm{o}}(w), \tag{A8}$$

where $\nabla J_{\mathrm{b}}(w) = \boldsymbol{B}w$ and $\nabla J_{\mathrm{o}}(w) = -\boldsymbol{B}\boldsymbol{H}^{\mathrm{T}}\boldsymbol{R}^{-1}(d - \boldsymbol{H}\boldsymbol{B}w)$, are the gradients of the background and the observation term respectively. If $\boldsymbol{B}$ is Gaussian-type, the role of $\boldsymbol{B}$ in Eqs (A7) and (A8) can be modeled by a recursive filter mentioned in the first part of Appendix.

Assuming $\boldsymbol{R}$ is an identity matrix, $\nabla J_{\mathrm{o}}(w)$ in Eq. (A8) becomes

$$\nabla J_{\mathrm{o}}(w) = -\boldsymbol{B}\boldsymbol{H}^{\mathrm{T}}(d - \boldsymbol{H}\boldsymbol{B}w). \tag{A9}$$

Under the same assumption and following the same analysis process as depicted in Section 3.2, we can conclude that the values of $\boldsymbol{H}^{\mathrm{T}}(d - \boldsymbol{H}\boldsymbol{B}w)$ are spatially incoherent, being zero at grid points with no measurements but not for those observed ones. Note that $\nabla J_{\mathrm{o}}(w)$ in Eq. (A9) is equivalent to applying a recursive filer on $-\boldsymbol{H}^{\mathrm{T}}(d - \boldsymbol{H}\boldsymbol{B}w)$. To retrieve maximum details both from the background and the observational field, the filter parameter should be chosen as a small one. As a result, the sharp variation of $\boldsymbol{H}^{\mathrm{T}}(d - \boldsymbol{H}\boldsymbol{B}w)$ cannot be smoothed out and thus will directly lead to the incoherent distribution of $\nabla J_{\mathrm{o}}(w)$ and $\nabla J(w)$. That means, observational information cannot be effectively transmitted to data void regions, causing an observational loss in $\nabla J(w)$. Without this observational constraint, when used by a conventional minimization algorithm, the "physically" erroneous $\nabla J(w)$ will cause analysis in those regions completely approach background field. Although the "jump" of $\boldsymbol{H}^{\mathrm{T}}(d - \boldsymbol{H}\boldsymbol{B}w)$ can be filtered out with a large filter parameter so that data-void regions can be filled up by long waves of observations, the analysis is overly smoothed and details are missing. In a word, no single selection of the filter parameter can fit all circumstances. A small value cannot make observational signals effectively transmit to data-void regions while a large one may only yield a coarse analysis.

To get a well-behaved analysis, inhomogeneous and anisotropic filters can be used with their parameters being determined from

error statistics so that in some regions they will present smaller filter parameters, while in other regions (e.g., data-void regions) they will show larger filter parameters. However, the practical implementation is difficult because we will never accurately know the true states of physical variables. Also, the construction of such filters is usually too complicated (Purser et al., 2003b). Therefore, homogeneous recursive filters are still in wide use in many practical applications.

Even with a simple recursive filter as discussed in the first part of Appendix, there are still possibilities of achieving a better analysis. It is should be noted that the gradient formulated by Eq. (A8) itself is a filtering process equivalent to applying a recursive filter to $w - \boldsymbol{H}^{\mathrm{T}}\boldsymbol{R}^{-1}(d - \boldsymbol{HB}w)$. However, this filtering is too weak to overcome a spatial incoherence with a small filter parameter when data are highly irregular distributed. Therefore, an extra filtering procedure can be conducted to enhance that filtering effect, just as what we have done in the SMRF scheme by filtering the gradient with a recursive filter at each iteration. Also, the filter parameter can decrease sequentially with iterations to provide a multi-scale analysis. This shows a potentiality of our objective analysis scheme to be extended to a 3D-VAR. However, further studies are still needed.