Scale Utilization and Optimization from Wavelet Analysis for Data Assimilation: SUgOiWADAi

LELAND JAMESON

Frontier Research System for Global Change and International Pacific Research Center, University of Hawaii at Manoa, Honolulu, Hawaii, and Hydrodynamics and Turbulence Theory and Simulation Group, University of California, Lawrence Livermore National Laboratory, Livermore, California

TAKUJI WASEDA AND HUMIO MITSUDERA

Frontier Research System for Global Change and International Pacific Research Center, University of Hawaii at Manoa, Honolulu, Hawaii

(Manuscript received 22 August 2000, in final form 12 July 2001)

ABSTRACT

The second generation of a new approach to data assimilation where wavelet analysis is used for error estimation is presented here. The first generation is known as EEWADAi. This modified and optimized method uses wavelet analysis to not only estimate numerical error but to also acquire an estimate of the variation at various scales of the model simulation. In the original EEWADAi, wavelet analysis on the finest scale was used to estimate numerical error. In the second-generation version, called SUgOiWADAi, wavelet analysis is used on a variety of scales to not only obtain an estimate of numerical error, finest-scale information, but to also obtain an estimate of model variation, information from coarser scales. This new algorithm is computationally very inexpensive and is very effective.

1. Introduction

In the field of data assimilation the key to success is knowing the errors in either the data or the computed values as a function of time and spatial location. It is rare that one will know such spatial and time information about the errors in the data, but wavelet analysis can give us such information on the errors of the computed values and hence a mechanism to determine an appropriate weighting between the data and computed values. Furthermore, wavelet analysis can give a very reliable estimate of model variation that often has a correlation with model error. This technique of correlating model error with model variation is used in the method known as optimal interpolation (OI; Ghil 1989).

In order to apply the so-called Kalman filter (Kalman 1960) to the meteorological and oceanographic application for data assimilation purposes, an approximation of the prediction error covariance matrix is necessary. Otherwise it would require the square of the model state dimension storage, $O(n^2)$, and $O(2n)$ additional integration time steps for the computation of the error evolution. Even with a sophisticated reduction method, initialization of the error covariance matrix remains a difficult task. Initialization of the prediction error is typically done by making an estimate based on, for example, a limited number of empirical orthogonal function (EOF) modes of the model variation (Pham et al. 1998).

In the new approach introduced by Jameson and Waseda (2000), the estimation of the error associated with numerical schemes was diagnosed by wavelet analysis and the detected error was used as an indicator of the prediction error. Therefore, it was not necessary to solve the evolution of the prediction error. Such a method can easily be implemented by extending a simple assimilation scheme often referred to as nudging (Malanotte-Rizzoli and Holland 1986) or OI (Mellor and Ezer 1991).

As an extension of the previous approach (EEWADAi) using the finest-scale wavelet coefficients that are known to detect numerical errors (Jameson 1998), we used in this study the wavelet coefficients on different scales that are considered to indicate model variation not necessarily restricted to numerical errors. In section

2, a brief review of wavelet diagnosis is given. Section 3 explains how one can measure variation at different scales and in multiple dimensions using wavelet decomposition. A short description of the assimilation scheme and the result of the twin experiment conducted using a regional ocean general circulation model (OGCM) will be presented in section 4, and the conclusion follows.

2. Wavelet analysis

Possibly the most instructive way to think of wavelets is in contrast to traditional analysis techniques such as Fourier analysis. With Fourier analysis we analyze discrete or continuous data using basis functions that are global, smooth, and periodic. This analysis yields a set of coefficients, say, a_k, which gives the amount of energy in the data at frequency k. Wavelet analysis, by contrast, analyzes data with basis functions that are local, slightly smooth, not periodic, and that vary with respect to scale and location. Wavelet analysis thereby produces a set of coefficients $b_{j,k}$ that give the amount of energy in the data at scale j and location k. Wavelet analysis can serve as a good complement to Fourier analysis. In fact, data that are efficiently analyzed with Fourier analysis often are not efficiently analyzed with wavelet analysis and the opposite situation also holds.

For our purposes here we will confine our discussion to the so-called orthogonal wavelets and specifically the Daubechies family of wavelets. The orthogonality property leads to a clear indication when data deviate from a low-order polynomial, the importance of which will become clear when we discuss numerical methods.

a. Defining the Daubechies wavelet

To define Daubechies-based wavelets, see Daubechies (1988) and Erlebacher et al. (1996), consider the two functions $\phi(x)$, the scaling function, and $\psi(x)$, the wavelet. The scaling function is the solution of the dilation equation:

$$\phi(x) = \sqrt{2} \sum_{k=0}^{L-1} h_k \phi(2x - k), \qquad (1)$$

where the coefficients h_k define the fundamental properties of of the scaling function and will be explained precisely in section 2c. Equation (1) carries the name "dilation equation" since the independent variable x appears alone on the left-hand side but is multiplied by 2, or dilated, on the right-hand side. One also requires the scaling function $\phi(x)$ be normalized: $\int_{-\infty}^{\infty} \phi(x)\, dx = 1$. The wavelet $\psi(x)$ is defined in terms of the scaling function:

$$\psi(x) = \sqrt{2} \sum_{k=0}^{L-1} g_k \phi(2x - k), \qquad (2)$$

see section 2c for an explanation of g_k.

One builds an orthonormal basis from $\phi(x)$ and $\psi(x)$ by dilating and translating to get the following functions:

$$\phi_k^j(x) = 2^{-j/2}\phi(2^{-j}x - k) \quad \text{and} \qquad (3)$$

$$\psi_k^j(x) = 2^{-j/2}\psi(2^{-j}x - k), \qquad (4)$$

where $j, k \in Z$. The dilation parameter is j, and k is the translation parameter.

b. The spaces spanned by wavelets

It is usual to let the spaces spanned by $\phi_k^j(x)$ and $\psi_k^j(x)$ over the parameter k, with j fixed, be denoted by $\mathbf{V}_j$ and $\mathbf{W}_j$ respectively,

$$\mathbf{V}_j = \sum_{k \in Z}^{\text{span}} \phi_k^j(x), \qquad (5)$$

$$\mathbf{W}_j = \sum_{k \in Z}^{\text{span}} \psi_k^j(x). \qquad (6)$$

The spaces $\mathbf{V}_j$ and $\mathbf{W}_j$ are related by

$$\cdots \subset \mathbf{V}_1 \subset \mathbf{V}_0 \subset \mathbf{V}_{-1} \subset \cdots, \quad \text{and} \qquad (7)$$

$$\mathbf{V}_j = \mathbf{V}_{j+1} \oplus \mathbf{W}_{j+1}, \qquad (8)$$

where the notation $\mathbf{V}_0 = \mathbf{V}_1 \oplus \mathbf{W}_1$ indicates that the vectors in $\mathbf{V}_1$ are orthogonal to the vectors in $\mathbf{W}_1$ and the space $\mathbf{V}_0$ is simply decomposed into these two component subspaces.

c. The high- and low-pass filters and orthogonality

The coefficients $H = \{h_k\}_{k=0}^{L-1}$ and $G = \{g_k\}_{k=0}^{L-1}$ are related by $g_k = (-1)^k h_{L-k}$ for $k = 0, \ldots, L - 1$. All wavelet properties are specified through the parameters H and G. If one's data are defined on a continuous domain such as $f(x)$ where $x \in R$ is a real number, then one uses $\phi_k^j(x)$ and $\psi_k^j(x)$ to perform the wavelet analysis. If, on the other hand, one's data are defined on a discrete domain such as $f(i)$ where $i \in Z$ is an integer then the data are analyzed, or filtered, with the coefficients H and G. In either case, the scaling function $\phi(x)$ and its defining coefficients H detect localized low-frequency information, that is, they are low-pass filters (LPF), and the wavelet $\psi(x)$ and its defining coefficients G detect localized high-frequency information, that is, they are high-pass filters (HPFs). Specifically, H and G are chosen so that dilations and translations of the wavelet, $\psi_k^j(x)$, form an orthonormal basis of $L^2(R)$ and so that $\psi(x)$ has M vanishing moments that determine the accuracy. In other words, $\psi_k^j(x)$ will satisfy

$$\delta_{kl}\delta_{jm} = \int_{-\infty}^{\infty} \psi_k^j(x)\psi_l^m(x)\, dx, \qquad (9)$$

where δ_{kl} is the Kronecker delta function, and the accuracy is specified by requiring that $\psi(x) = \psi_0^0(x)$ satisfy

$$\int_{-\infty}^{\infty} \psi(x)x^m \, dx = 0, \qquad (10)$$

for $m = 0, \ldots, M - 1$. Under the conditions of the previous two equations, for any function $f(x) \in L^2(R)$ there exists a set $\{d_{jk}\}$ such that

$$f(x) = \sum_{j \in Z} \sum_{k \in Z} d_k^j \psi_k^j(x), \qquad (11)$$

where

$$d_k^j = \int_{-\infty}^{\infty} f(x)\psi_k^j(x) \, dx. \qquad (12)$$

d. Quadrature mirror filters and the Haar wavelet

The two sets of coefficients H and G are known as quadrature mirror filters. For Daubechies wavelets the number of coefficients in H and G, or the length of the filters H and G, denoted by L, is related to the number of vanishing moments M by $2M = L$. For example, the famous Haar wavelet is found by defining H as $h_0 = h_1 = 1$. For this filter, H, the solution to the dilation equation, Eq. (1); $\phi(x)$, is the box function: $\phi(x) = 1$ for $x \in [0, 1]$ and $\phi(x) = 0$ otherwise. The Haar function is very useful as a learning tool, but because of its low order of approximation accuracy and lack of differentiability it is of limited use as a basis set. The coefficients H needed to define compactly supported wavelets with a higher degree of regularity can be found in Daubechies (1988). As expected, the regularity increases with the support of the wavelet. The usual notation to denote a Daubechies-based wavelet defined by coefficients H of length L is D_L.

e. Setting a largest and smallest scale

In a continuous wavelet expansion, functions with arbitrarily small-scale structures can be represented. In practice, however, there is a limit to how small the smallest structure can be, depending on, for example, the numerical grid resolution or the sampling frequency in a signal processing scenario. Hence, on a computer an approximation would be constructed in a finite space such as

$$\mathbf{V}_0 = \mathbf{W}_1 \oplus \mathbf{W}_2 \oplus \cdots \oplus \mathbf{W}_J \oplus \mathbf{V}_J,$$

with the approximation being

$$P_{V_0} f(x) = \sum_{k \in Z} s_k^J \phi_k^J(x) + \sum_{j=1}^{J} \sum_{k \in Z} d_k^j \psi_k^j(x), \qquad (13)$$

with

$$d_k^j = \int_{-\infty}^{\infty} f(x)\psi_k^j(x) \, dx, \qquad s_k^J = \int_{-\infty}^{\infty} f(x)\phi_k^J(x) \, dx$$

utilizing orthogonality. Within this expansion, the scale $j = 0$ is arbitrarily chosen as the finest scale required, and scale J would be the scale at which a kind of local average, $\phi_k^J(x)$, provides sufficient large-scale information, that is, the first term in Eq. (13) provides the local mean around which the function oscillates.

One must also limit the range of the location parameter, k. Assuming periodicity of $f(x)$ implies periodicity on all wavelet coefficients, s_k^j and d_k^j, with respect to k. For the nonperiodic case, since k is directly related to the location, a limit is imposed on the values of k when the location being addressed extends beyond the boundaries of the domain.

f. Implementation on a computer

The wavelet decomposition matrix is the matrix embodiment of the dilation equation, Eq. (1), defining the scaling function and the accompanying equation defining the wavelet, Eq. (2). The following two recurrence relations for the coefficients, s_k^j and d_k^j, in Eq. (13) are given as

$$s_k^j = \sum_{n=1}^{L} h_n s_{n+2k-2}^{j-1} \quad \text{and} \qquad (14)$$

$$d_k^j = \sum_{n=1}^{L} g_n s_{n+2k-2}^{j-1} \qquad (15)$$

as obtained from Eqs. (1)–(2), and we recall that h_n refers to the chosen filter while we have $g_n = -(-1)^n h_{L-n}$.

Denote the decomposition matrix embodied by these two equations, assuming periodicity, by $\mathbf{P}_N^{j,j+1}$ where the matrix subscript denotes the size of the square matrix while the superscripts indicate that $\mathbf{P}$ is decomposing from scaling function coefficients at scale j to scaling function and wavelet function coefficients at scale $j + 1$, that is, $\mathbf{P}_N^{j,j+1}$ maps $\mathbf{s}_j$ onto $\mathbf{s}_{j+1}$ and $\mathbf{d}_{j+1}$:

$$\mathbf{P}_N^{j,j+1}: \quad [\mathbf{s}_j] \rightarrow \begin{bmatrix} \mathbf{s}_{j+1} \\ \mathbf{d}_{j+1} \end{bmatrix}, \qquad (16)$$

where we by $\mathbf{s}_j$ refer to the vector containing the coefficients at scale j. Note that the vectors at scale $j + 1$ are half as long as the vectors at scale j.

To be perfectly correct one would first approximate the scaling function coefficients at the finest scale using the raw data, however, in practice it seems to make very little difference if one simply considers the raw data to be the scaling function coefficients. So, for our purposes here we will simply use the raw data as the scaling function coefficients on the finest scale. The repeated application of the matrix $\mathbf{P}^{j,j+1}$ yields the wavelet coefficients at the various scales, and it is these wavelet coefficients that provide a guide to the errors committed during the numerical calculation.

To illustrate further, let us consider that the raw data are given and are assumed to be the scaling function coefficients on the finest scale, $\mathbf{s}_0$. One wavelet decom-

position yields the scaling function coefficients and wavelet coefficients at scale $j = 1$, $\mathbf{s}_1$, and $\mathbf{d}_1$. A second application of the wavelet decomposition matrix will yield the vectors $\mathbf{s}_2$ and $\mathbf{d}_2$. It is the vectors $\mathbf{d}_1$, $\mathbf{d}_2$, . . . , which yield the critical information on the numerical errors. If, for example, one sees that the values of the $\mathbf{d}_1$ are relatively large in the middle of the vector, then it is clear that within this one-dimensional vector the largest errors will be in the middle of the one-dimensional domain from which this vector was derived. What we care about most are the relative errors being committed, but we also have some interest in the absolute errors, the subject of the next section.

3. SUgOiWADAi

As we mentioned in the abstract and introduction, SUgOiWADAi works by using the information from many scales of the wavelet analysis in contrast to EEWADAi that used the wavelet analysis information only on the finest scale. By increasing the amount of information used, one would naturally expect that the result would improve. In particular, SUgOiWADAi provides not only a measure of numerical error as EEWADAi does, but it also provides an estimate of model variation. In order to understand how this variation is measured and how this new information is coupled with the estimate of numerical error, we need to return to the wavelet subspace notation.

a. Measuring variation at different scales in one dimension

As we have mentioned, SUgOiWADAi not only can give a direct measure of numerical error but the scheme can also give a very reliable measure of model variation. In this section we will explain how SUgOiWADAi works and compare the variation measure given by SUgOiWADAi to that used in OI schemes.

From wavelet analysis one can find the variation at a variety of scales. We can think of the information in the highest-frequency wavelet box $\mathbf{W}_1$ as telling us about the numerical error, that is, it will give us a measure of local deviation from low-order polynomials. It can also be seen as a measure of variation at the smallest scale or similarly the highest frequency. Basically, the information in the box $\mathbf{W}_1$ measures variation over roughly two grid points. In order to compare the ability of wavelets to detect variation with the ability of OI schemes to detect variation, we need a measure of variation that covers a slightly larger portion of the domain. Therefore, we would use the information available in the boxes $\mathbf{W}_2$, $\mathbf{W}_3$, and perhaps also $\mathbf{W}_4$, where the box numbering is such that higher numbers represent coarser scales.

1) THE HAAR (D2) WAVELET

First we will show concretely how the Haar wavelet measures variation as we go from fine to coarse wavelet subspaces. First note that the Haar wavelet filter has only two nonzero coefficients. We will denote the low-pass filter coefficients by h_0 and h_1 and the high-pass filter coefficients by g_0 and g_1. So, if we are given a stream of data from one field, perhaps velocity, pressure, etc., from our calculation, say, f_0, f_1, . . . , f_N, then the first step is to find the scaling function coefficients on the finest scale, s_0^0, s_1^0, . . . , s_N^0 from this data stream. Recall that these finest-scale coefficients are the coefficients corresponding to the scaling function subspace $\mathbf{V}_0$. Stricktly speaking, one should approximate the scaling function coefficients on the finest scale from the raw data via some kind of quadrature formula; however, our goal is to measure only the variation in the data, therefore we can simply assign these scaling functions coefficients directly to the raw data: $s_0^0 = f_0$, $s_1^0 = f_1$, . . . , $s_2^0 = f_2$. Then one obtains the ''scaling'' function coefficients at the first scale, s_0^1, s_1^1, . . . , $S_{N/2}^1$, that is, the coefficients for the box $\mathbf{V}_1$, and the ''wavelet'' coefficients d_0^1, d_1^1, . . . , $d_{N/2}^1$, that is, the coefficients for the box $\mathbf{W}_1$, by

$$s_0^1 = (h_0 s_0^0) + (h_1 s_1^0) \qquad s_1^1 = (h_0 s_2^0) + (h_1 s_3^0) \quad \text{and}$$

$$d_0^1 = (g_0 s_0^0) + (g_1 s_1^0) \qquad d_1^1 = (g_0 s_2^0) + (g_1 s_3^0).$$

The important point to note here is that the wavelet coefficients on this first scale are combined from only two scaling function coefficients on the finest scale, therefore they can only *feel* variation over only two numbers. Now, to go from this first scale to the second scale,

$$\mathbf{V}_1 = \mathbf{V}_2 \oplus \mathbf{W}_2,$$

then, similarly, each coefficient in the subspaces $\mathbf{V}_2$ and $\mathbf{W}_2$ is found from two numbers in the subspace $\mathbf{V}_1$. Therefore, each coefficient in $\mathbf{W}_2$ is composed from four numbers in the finest-scale subpace $\mathbf{V}_0$. Therefore, one can say that each coefficient in $\mathbf{W}_2$ can *feel* or measure variation that occurs over four numbers in the original raw data f_i. Likewise, each wavelet coefficient in $\mathbf{W}_3$ can *feel* variation that occurs over eight numbers in the original data.

It is very important to note that the Haar wavelet is very special in the sense that the wavelet coefficients do not overlap when performing a wavelet decomposition. More will be said on this point in the next section.

2) THE D4 WAVELET

As one can see from Eqs. (14) and (15), when the wavelet filtering algorithm is applied there is overlap of neighboring wavelet filtering coefficients. This is clearly seen when the wavelet decomposition is put in matrix form. For the 4-coefficient $\mathbf{D}_4$ wavelet, suppose that one wants to project from 8 scaling function coefficients at scale j to 4 scaling function coefficients at scale $j + 1$ and 4 wavelet coefficients at scale $j + 1$. The decomposition matrix for the case of periodic boundary conditions, $\mathbf{P}_8^{j,j+1}$, thus becomes

$$\mathbf{P}_8^{j,j+1} \equiv \begin{bmatrix} h_1 & h_2 & h_3 & h_4 & 0 & 0 & 0 & 0 \\ 0 & 0 & h_1 & h_2 & h_3 & h_4 & 0 & 0 \\ 0 & 0 & 0 & 0 & h_1 & h_2 & h_3 & h_4 \\ h_3 & h_4 & 0 & 0 & 0 & 0 & h_1 & h_2 \\ g_1 & g_2 & g_3 & g_4 & 0 & 0 & 0 & 0 \\ 0 & 0 & g_1 & g_2 & g_3 & g_4 & 0 & 0 \\ 0 & 0 & 0 & 0 & g_1 & g_2 & g_3 & g_4 \\ g_3 & g_4 & 0 & 0 & 0 & 0 & g_1 & g_2 \end{bmatrix}, \quad (17)$$

where the periodicity is reflected in the coefficients being wrapped around. One can see in each row of the above matrix that the coefficients are simply shifted by 2 as one goes from the upper rows to the lower rows. Given this information, let us examine the region that can be *felt* by each coefficient in the various wavelet subspaces $\mathbf{W}_1$, $\mathbf{W}_2$, and $\mathbf{W}_3$. Obviously, as we go from $\mathbf{V}_0$ to $\mathbf{V}_1 \oplus \mathbf{W}_1$ we can see that each wavelet coefficient d_1^k in $\mathbf{W}_1$ feels a region of four numbers in the original data.

Now, let us examine the region that can be felt in $\mathbf{W}_2$. In a straightforward application of the above matrix, we can see that in our first step in going from $\mathbf{V}_0$ to $\mathbf{V}_1$ we have

$$s_1^1 = (h_1 f_0) + (h_2 f_1) + (h_3 f_2) + (h_4 f_3)$$

$$s_1^2 = (h_1 f_2) + (h_2 f_3) + (h_3 f_4) + (h_4 f_5)$$

$$s_1^3 = (h_1 f_4) + (h_2 f_5) + (h_3 f_6) + (h_4 f_7)$$

$$s_1^4 = (h_1 f_6) + (h_2 f_7) + (h_3 f_8) + (h_4 f_9)$$

and when we decompose $\mathbf{V}_1$ into $\mathbf{V}_2 \oplus \mathbf{W}_2$ we see that

$$d_2^1 = (g_1 s_1^1) + (g_2 s_1^2) + (g_3 s_1^3) + (g_4 s_1^4.)$$

Now, we can see that the region that can be felt in the coefficient d_2^1 extends from f_0 to f_9, that is, a region 10 points wide. Thus as we go to wavelet and scaling function coefficients at higher and higher levels of decomposition, one can see that the coefficients are influenced by more and more of the data in the physical space, or the coefficients feel an increasingly larger region of the physical space. So, after one decomposition, each coefficient will feel a region equal in size to the length of the wavelet filter denoted by the parameter L. And, as one decomposes on higher and coarser scales the region felt by each wavelet coefficient grows in proportion to the length of the wavelet filter and the level of decomposition.

b. Wavelet detection of model variation

As explained in previous sections of the paper, wavelet analysis breaks up data into local frequency components. That is, at a given physical space location one can obtain an estimate of the various scales of information present in the vicinity of this physical space location. Using the previously defined notation, we have
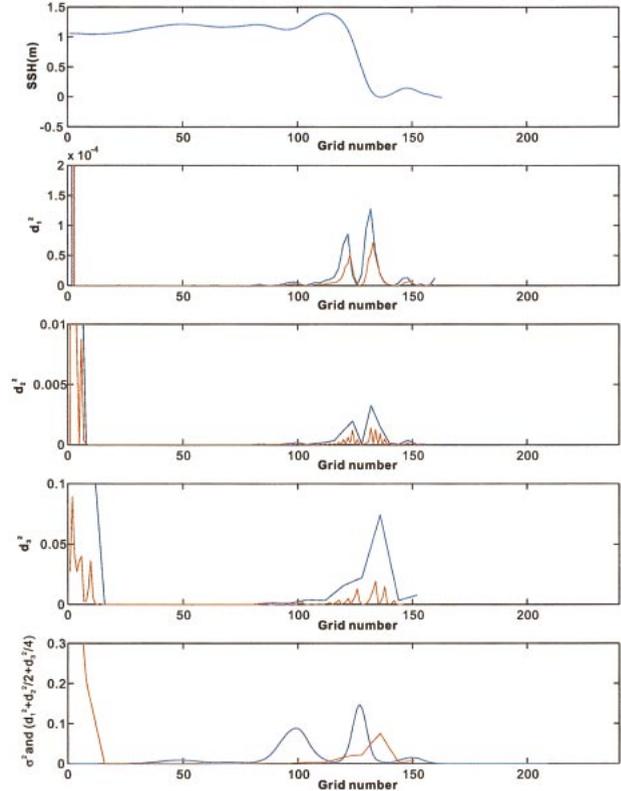

FIG. 1. Original SSH profile that crosses the Kuroshio Current is plotted in the top panel and the information from wavelet analysis vs grid number at three scales are shown below. Blue curves in the second, third, and fourth panels show the wavelet coefficients from finest to coarser scales (d_1, d_2, and d_3), and partial reconstruction at each scale are shown in red curves. Combination of the wavelet coefficients ($d_1^2 + d_2^2/2 + d_3^2/4$) are plotted by a red curve at the bottom panel together with the blue curve indicating the averaged model variation (σ^2).

the "frequency" or "scale" boxes $\mathbf{W}_1$, $\mathbf{W}_2$, $\mathbf{W}_3$, . . . , which contain this local frequency content information. Furthermore, "variation" in a model will appear as a localized oscillation at some scale, and this information will appear as a large wavelet coefficient in one of the frequency boxes. That is, there is a one-to-one correspondence between model variation at a given scale and the wavelet energy at that same scale. In fact, variation is exactly what wavelet analysis detects. One can see these effects in Fig. 1. In the top panel of the figure, one can see the sea surface height (SSH) as one traverses the Kuroshio. In the second panel from the top, one can see the finest-scale wavelet coefficient. This coefficient has two peaks according to the location where the numerical truncation error will be the largest. The third and fourth panels show the wavelet coefficients at the larger scale of the second and third decompositions. At these scales one is detecting not numerical error but model variation.

In summary, we note that other techniques such as OI measure only model variation, and this measure of

variation is far less precise than the model variation measure that is given by wavelet analysis. In addition, wavelet analysis gives a precise measure of the error committed in the numerical calculation. Both of these estimates, the error estimate and the variation estimate, make wavelet analysis a very powerful and useful tool.

c. When variation and error are not the same

In a word, where there is *numerical error* there will be variation, but variation does not imply numerical error. Roughly speaking, one can make the argument based on scale information. For example, if one observes wavelet energy in the finest-scale box, $\mathbf{W}_1$, then this energy will indicate that the gridpoint density for the numerical method in a given region of the domain is not sufficient and that numerical errors are committed in this region. Certainly this same $\mathbf{W}_1$ box energy indicates local high-frequency information or local variation. On the other hand, observed energy in the coarser-scale box, $\mathbf{W}_3$, will not necessarily indicate an insufficient gridpoint density but, again, it will indicate that variation occurs at scale 3. To be more precise, suppose that in the vicinity of gridpoint x_k that energy is present in box $\mathbf{W}_3$ but not in box $\mathbf{W}_1$, this will indicate variation but not numerical error. On the other hand, if energy is present in box $\mathbf{W}_1$ then this will indicate both numerical error and variation. As above, let us refer to Fig. 1 where we can see that the third and fourth panels from the top show the model variarion but not the numerical truncation error. At these scales the model physics are producing changes or variations but the scale is sufficiently large that the numerical truncation error is very small.

d. Measuring variation at different scales in higher dimensions

Measuring the variation in a given field in higher dimensions is a straightforward extension of the ideas from one dimension. One can simply perform the wavelet analysis in a tensor product approach dimension by dimension. In fact, the regions of the domain that can be *felt,* as discussed above, carry over directly dimension by dimension. The tensor product approach works as follows: recall that in one dimension, one decomposition of the finest-scale subspace yields

$$\mathbf{V}_0 = \mathbf{V}_1 \oplus \mathbf{W}_1.$$

So, if one takes the tensor product of this one-dimensional analysis with another one-dimensional analysis then one obtains

$$\mathbf{V}_0 \otimes \mathbf{V}_0 = (\mathbf{V}_1 \oplus \mathbf{W}_1) \otimes (\mathbf{V}_1 \oplus \mathbf{W}_1), \qquad (18)$$

which yields

$$\mathbf{V}_0 \otimes \mathbf{V}_0 = (\mathbf{V}_1 \otimes \mathbf{V}_1) \oplus (\mathbf{V}_1 \otimes \mathbf{W}_1)$$
$$\oplus (\mathbf{W}_1 \otimes \mathbf{V}_1) \oplus (\mathbf{W}_1 \otimes \mathbf{W}_1). \qquad (19)$$

One can think of the subspace,

$$\mathbf{V}_1 \otimes \mathbf{V}_1$$

as representing the horizontal and vertical average, low-pass filtering, of the information contained in $\mathbf{V}_0 \otimes \mathbf{V}_0$. The subspace

$$\mathbf{V}_1 \otimes \mathbf{W}_1$$

represents a horizontal low-pass-filtering process and a vertical high-pass-filtering process. Such a subspace would capture horizontal edges. On the other hand, the subspace

$$\mathbf{W}_1 \otimes \mathbf{V}_1$$

would represent a horizontal high-pass-filtering process and a vertical low-pass-filtering process. One would be able to detect vertical edges in such a subspace. Finally, when the high-pass filter is applied in both the vertical and horizontal directions one arrives at the subspace,

$$\mathbf{W}_1 \otimes \mathbf{W}_1.$$

It is this subspace that we are primarily interested in since it will detect variation in both the horizontal and vertical directions. Note that as in the one-dimensional case we will certainly perform many levels of wavelet decompositions, and the subspaces that we will primarily be interested in will be

$$\mathbf{W}_j \otimes \mathbf{W}_j$$

for $j = 1, 2, 3, 4..,$. Generally, decompositions up to $j = 4$ will be sufficient.

NUMERICAL ERROR AND VARIATION IN HIGHER DIMENSIONS

We note that it is the subspace $\mathbf{W}_1 \otimes \mathbf{W}_1$ that will primarily detect numerical error. Recall that the reason for this is that this subspace will detect deviation from low-order polynomials over just a few grid points in both the horizontal and vertical directions. In Fig. 2 one can see the four subspaces formed from the two-dimensional wavelet decomposition at the finest scale. The upper-left panel contains the wavelet coefficients at the finest scale in both directions. This panel would indicate the presence of numerical truncation error in both directions simultaneously. The upper-right panel of Fig. 2 indicates the presence of numerical truncation error in the vertical direction, and the lower-left panel indicates the presence of numerical truncation error in the horizontal direction. The lower-right panel is a simultaneous average in both directions; note the change of color scale in this panel. Figure 3 indicates the two-dimensional wavelet coefficients after a second wavelet decomposition and Fig. 4 indicates the two-dimensional wavelet coefficients after a third decomposition. Figures 3 and 4 would be used to give estimates of model variation. Further, a subspace such as $\mathbf{W}_4 \otimes \mathbf{W}_4$ will detect deviation over a very large number of grid points in both the horizontal and vertical directions. Note that in
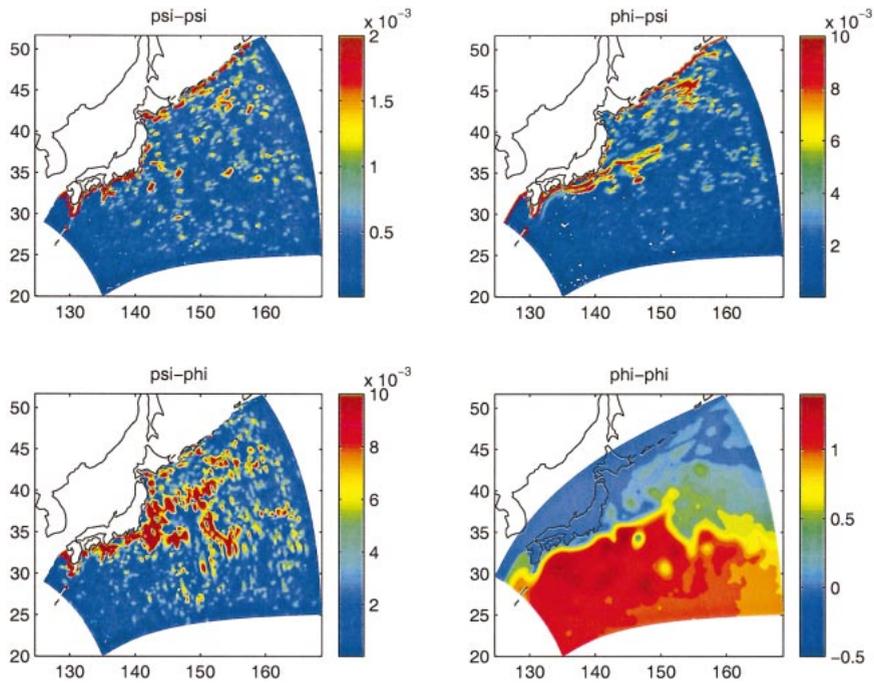
FIG. 2. Magnitude of the wavelet coefficients at the finest wavelet scale in the four wavelet boxes defined in the text. Such finescale information is a reliable detector of numerical error.
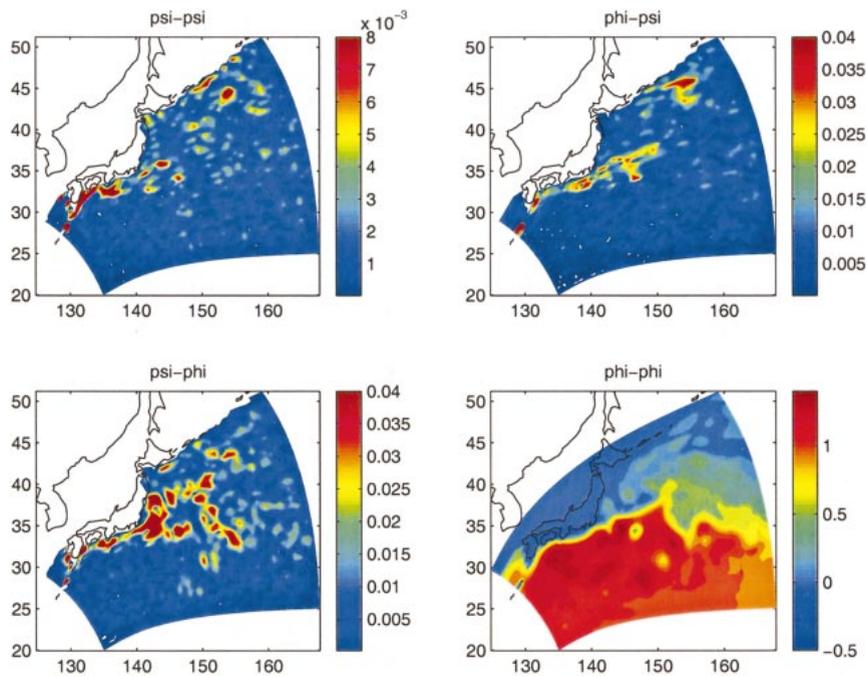
FIG. 3. Magnitude of the wavelet coefficients at scale 2 in the 4 wavelet boxes. Such information at this scale is a good indicator of model variation at finescale.
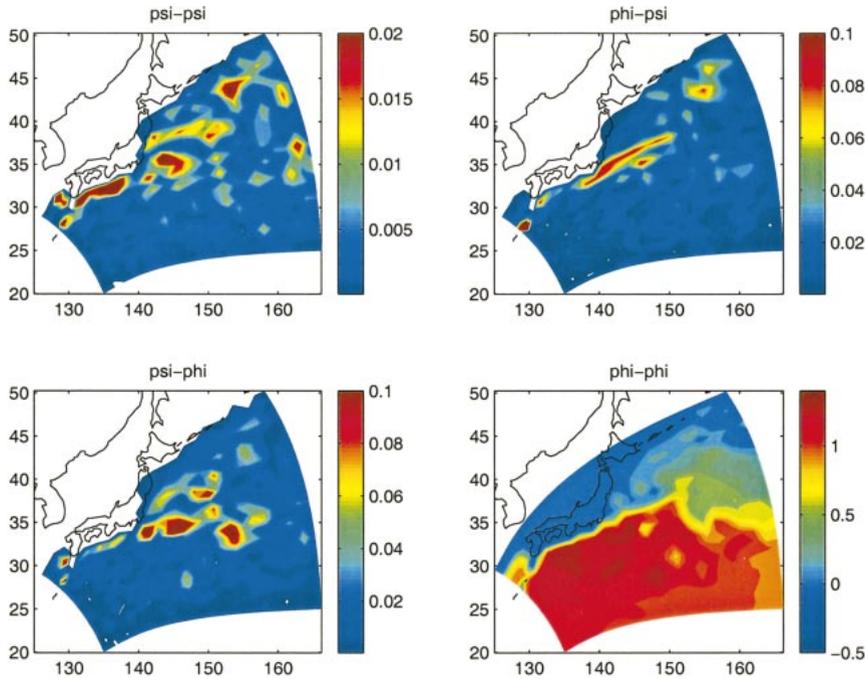
FIG. 4. Magnitude of the wavelet coefficients at scale 3 in the 4 wavelet boxes. Such information at this relatively coarse scale is a reliable indicator of model variation.

Figs. 2, 3, and 4 the lower-right panel denotes the values of the scaling function coefficients. These lower-right panels give, if you will, a local average of the data and should look essentially like a smeared version of the original data. In Fig. 5 we show the wavelet coefficients that combine the numerical error with the model variation. As mentioned above, deviation from low-order polynomials over a large number of grid points *does not* imply numerical error but simply gives a measure of variation. As one proceeds down the heirarchy of subspaces from $\mathbf{W}_1 \otimes \mathbf{W}_1$ to $\mathbf{W}_4 \otimes \mathbf{W}_4$ then one is proceeding from a very localized measure of deviation to a more global measure of deviation. It is only the most localized measure of deviation that can provide a measure of numerical error, whereas the the more global measures of deviation can be used to give a measure of variation.

As above, we can refer to Fig. 1 to see an illustration of these ideas. In terms of the tensor product, the explanation is simple since one can consider the information direction by direction. In this case, the explanation is exactly the same as in the one-dimensional case if one considers the explanation direction by direction. Thus, for one to see large wavelet coefficients in the subspace $\mathbf{W}_1 \otimes \mathbf{W}_1$ then one must see large wavelet coefficients in each direction of the first scale decomposition, which is the second panel of Fig. 1. On the other hand, large wavelet coefficients in the subspace $\mathbf{W}_3 \otimes \mathbf{W}_3$ would require large wavelet coefficients in each direction of the third level of decomposition corresponding to the fourth panel of Fig. 1.

e. error variance estimation using wavelet analysis

The key difference between EEWADAi and SUg-OiWADAi is that SUgOiWADAi uses wavelet information at a variety of scales in order to get an estimate of not only *numerical* variation as is done in EEWADAi, but also an estimate of model variation, similar to what is done in OI methods. Recall that in EEWADAi wavelet analysis gives us a local measure of deviation from the low-order polynomial structure in the data. Given this measure of deviation, then one has an estimate of local numerical error in the model. From these local error estimates, EEWADAi easily provides an estimate of local error variance. For EEWADAi it was important that this variance estimate remain *local* since the ability to estimate errors locally in both space and time was one of the key strong points of EEWADAi. Recall that for EEWADAi we found a local average of the squared errors in space and we estimated the error variance around, say, grid point $(x_{K_1}, y_{K_2}, z_{K_3})$ in the following manner:

$$\sigma^2_{K_1, K_2, K_3} = \frac{1}{(2n+1)^3} \sum_{k_1=K_1-n}^{k_1=K_1+n} \sum_{k_2=K_2-n}^{k_2=K_2+n} \sum_{k_3=K_3-n}^{k_3=K_3+n} Cd(k_1, k_2, k_3)^2,$$

(20)

where $d(k_1, k_2, k_3)$ is the wavelet coefficient after only one decomposition at the wavelet spatial index of k_1, k_2, k_3. Here n is a small number used to define a small averaging box around the point of interest, and C is a constant that is needed to scale the wavelet transform to the problem at hand.
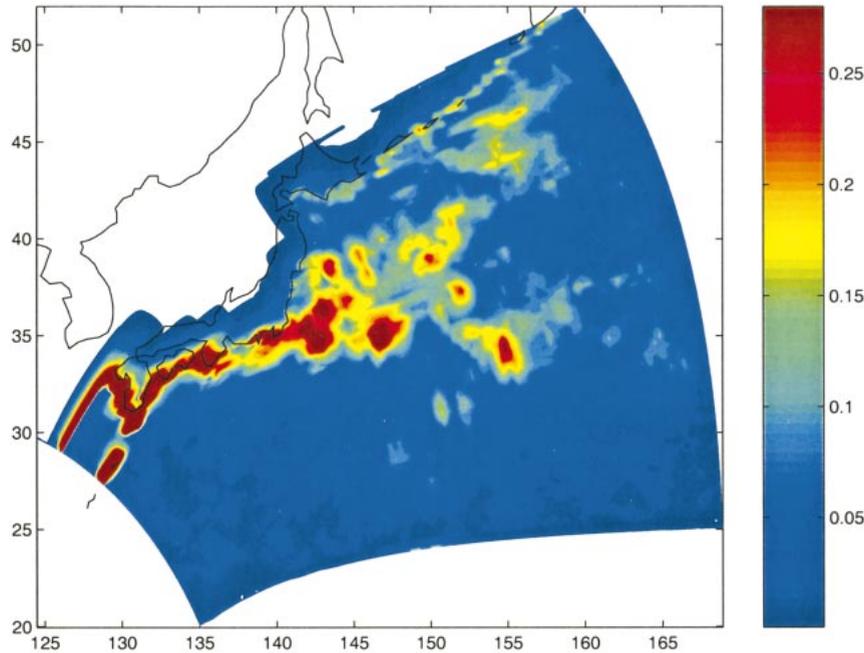
FIG. 5. Magnitude of the wavelet coefficients from wavelet analysis at three scales combined to yield an error covariance matrix that incorporates both numerical error and model variation.

For the case of SUgOiWADAi the averaging process is a little more complicated and one obtains

$$\sigma^2_{K_1,K_2,K_3} = \sum_{j=1}^{j=J} C_j \Delta_j, \qquad (21)$$

where Δ_j will be the wavelet-detected variation at scale j, and C_j will be a corresponding scaling constant. We have chosen the coefficients C_j to correspond to a natural scaling that occurs within the wavelet basis. That is, we choose $C_1 = 1$, $C_2 = 1/2$, and $C_3 = 1/4$. It should be stated that some flexibility exists here and one might want to alter these coefficients depending the characteristics of the calculation at hand. For example, altering the order of the numerical scheme can vastly change the magnitude of the coefficients in the finest-scale subspace $\mathbf{W}_1$. In this case, one might choose to increase C_1 to a number scaled according to this numerical order. Again, the user should experiment with these constants to find values that are suitable for the given calculation; Δ_j is defined as

$$\Delta_j = \frac{1}{(2n_j + 1)^3} \sum_{k_1=K_{j1}-n_j}^{k_1=K_{j1}+n_j} \sum_{k_2=K_{j2}-n_j}^{k_2=K_{j2}+n_j} \sum_{k_3=K_{j3}-n_j}^{k_3=K_{j3}+n_j} [d_j(k_1, k_2, k_3)]^2.$$
$$(22)$$

As above, the parameter n_j will define a box in three dimensions about which the summation of the wavelet coefficients occurs. If one is working in a lower dimension such as two dimensions, then one would not sum over the the k_3 parameter. The reason that one would have a different n_j for each scale j is that the wavelets at the larger scales, higher values of j, will cover larger portions of the domain. One can see this from the above discussion of regions of influence. Therefore, one would expect that the values of n_j will decrease as j increases. This will, of course, depend on the size of the region that one wished to use in finding their estimate of error variance. The point (K_{j1}, K_{j2}, K_{j3}) will be a wavelet translation index at the scale j that will correspond to a point in the physical space that is roughly centered near the area of the physical space where one needs an estimate of error variance.

4. Comparison with optimal interpolation

In a word, optimal interpolation (OI) schemes work by using the model variation as an estimate of model error. In fact, it is easy to find counter examples that illustrate that the model variation can be independent of model error, but in practical computations there appears to be a correlation. Therefore, it is common practice to use the model variation as an estimate of model error in OI schemes.

a. Computational cost

One certain advantage that wavelet analysis will have over OI schemes is that the wavelet analysis can be performed at every time step at a very low computational cost. For example, if one is using the $\mathbf{D}_4$ wavelet then the computational cost will be $4 \times N$ to decompose once, which gives us the information for the $\mathbf{W}_1$ box,

and $4 \times N/2$ to get the information in the $\mathbf{W}_2$ box, leading to a total computational work of $4 \times N + 4 \times N/2 + 4 \times N/4 + 4 \times N/8$ to obtain the information in the $\mathbf{W}_4$ box. Computationally this is relatively inexpensive. If, however, one considers this to be too expensive, then one can certainly perform the analysis every other time step or even less often. This depends on how quickly the data changes with respect to a given time step. Usually the solution will not change much qualitatively within just a few time steps.

b. Measuring variation and variance

In Fig. 1 we show a comparison between the instantaneous measure of model variation and numerical error given by wavelet analysis and the time-averaged variance that is used in the method of OI as an estimate for model error. Note that the blue curves in plots 2, 3, and 4 are the squares of the wavelet coefficients at wavelet scales 1, 2, and 3. At scales 1 and 2 we can see that the blue curve has 2 peaks that correspond to the 2 regions of the top plot that contain the small scale, that is, the regions of the top plot where the curve bends the most. At scale 3 the blue curve has 1 large peak that captures the entire downward sloping region of the top plot in addition to the 2 curving regions before and after the downward slope. In other words, at wavelet scale 3 one sees a rather large-scale phenomenon. In the bottom plot the blue curve indicates the time-averaged variance that is used in OI methods as an indicator or model error. However, as one can clearly see from the peaks of this blue curve in the bottom plot, the time-averaging process tends to smear information, thereby producing a less accurate estimate of instantaneous model variation or error.

c. Wavelet reconstruction at different scales

In Fig. 1 we not only show the square of the wavelet coefficients at different scale but also the so-called wavelet ''reconstruction'' at that scale. For example, the sea surface height $h(x)$ might have a wavelet expansion of

$$h(x) = \sum_{k=1}^{N/2} d_{1,k}\psi_{1,k}(x) + \sum_{k=1}^{N/4} d_{2,k}\psi_{2,k}(x)$$
$$+ \sum_{k=1}^{N/8} d_{3,k}\psi_{3,k}(x) + \sum_{k=1}^{N/8} s_{3,k}\phi_{3,k}(x), \quad (23)$$

where the ds indicate the wavelet coefficients and the ss indicate the scaling function coefficients. In Fig. 1 we see that the blue curves in the second, the third and the fourth plots show the square of the ds, and the red curves in the same three figures show the partial reconstructions at the given scales. In other words, the red curve in the second plot shows

$$h_1(x) = \sum_{k=1}^{N/2} d_{1,k}\psi_{1,k}(x),$$

which would be the projection of $h(x)$ onto the wavelet basis functions at the finest scale. One will notice that whereas the blue curves are relatively smooth, the red curves are relatively rough. This roughness is due to the nature of the $\mathbf{D}_4$ wavelet that is currently being used.

d. Numerical setup for the twin experiment

Twin experiments were conducted in order to perform a benchmark test of the three schemes, OI, EEWADAi, and SUGOiWADAi. The numerical setup of the twin experiment is similar to the one described in Jameson and Waseda (2000) and so the details will be omitted here. It is a standard method of testing the convergence of the assimilated solution to the control run by means of comparing their square differences.

The model is a version of a sigma-coordinate primitive equation (Mitsudera et al. 1997) that covers the main Kuroshio stream along the southern Japan coast, the Oyashio Current in the north, and the Kuroshio Extension region; the approximate domain is from 20°–52°N to 125°–170°E, configured to have a curvilinear coordinate system (206 by 209 by 32 discretization) in which the horizontal axes follow the mean geometry of Kuroshio stream. The set of dynamical and thermodynamic primitive equations are described in Blumberg and Mellor (1983). During the 6 years of spinup integration, the model was driven by monthly mean climatologies such as Hellerman–Rosenstein wind and the Comprehensive Ocean–Atmosphere Data Set (COADS) heat flux, and the Levitus climatology was used for the temperature and salinity initialization as well as for the lateral boundary restoration.

The assimilation method is a standard reduced Kalman filtering scheme where the correction is made only for a reduced set of model variables; a statistical correlation is used to map the surface elevation correction to the interior temperature and salinity correction (e.g., Ezer and Mellor 1994). Further reduction is made by assuming a certain shape of a model error covariance matrix. Then the gain matrix $\mathbf{K}$ is obtained from the following set of equations:

$$\mathbf{K} = \mathbf{SH}^\mathsf{T}[\mathbf{HSH}^\mathsf{T} + \mathbf{R}]^{-1} \quad (24)$$

$$\mathbf{S} = (\mathbf{D})^{1/2}\mathbf{C}(\mathbf{D})^{1/2}, \quad (25)$$

where the approximate forecast error covariance matrix $\mathbf{S}$ is given as a product of the time-independent correlation matrix $\mathbf{C}$ and a diagonal variance matrix $\mathbf{D}$. In the current implementation, the matrix $\mathbf{C}$ is a Gaussian distance-dependent function:

$$C_{ij} = \exp(-\|\mathbf{r}_i - \mathbf{r}_j\|^2/s_o^2), \quad (26)$$

with the decorrelation length scale s_o of 60–100 km. The diagonal elements of matrix $\mathbf{D}$ were fixed for the
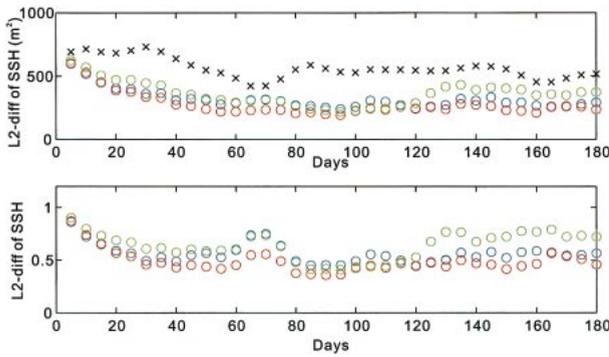
FIG. 6. (top) Convergence of the assimilation run (OI: green, EE-WADAi: blue, and SUgOiWADAi: red) to the control run showing the L2 difference of SSH vs days. The cross (×) is the L2 difference of the simulation run and the control run. (bottom) The same curves normalized by the L2 difference of the simulation run to the control run.
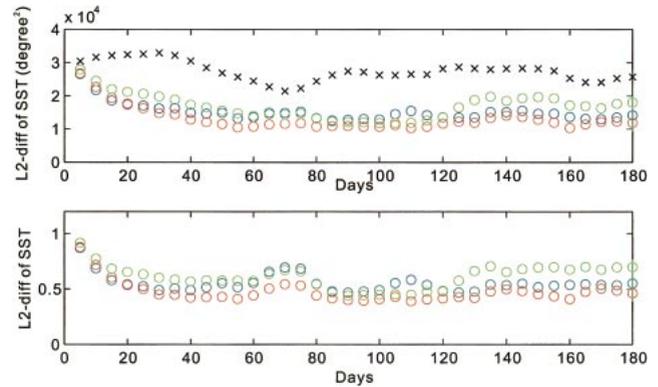
FIG. 7. (top) Convergence of the assimilation run (OI: green, EE-WADAi: blue, and SUgOiWADAi: red) to the control run showing the L2 difference of SST vs days. The cross (×) is the L2 difference of the simulation run and the control run. (bottom) The same curves normalized by the L2 difference of the simulation run to the control run.

OI scheme initialized by the model variance and were varied in time and space for EEWADAi and SUGOiWADAi. The elements of the diagonal observation error matrix $\mathbf{R}$ were fixed in time so the only change in $\|\mathbf{K}\|$ would occur through variation in $\mathbf{D}$. The matrix $\mathbf{H}$ represents mapping of the satellite data along the track onto the nearest neighboring model grid points. The gain matrix $\mathbf{K}$ is computed once for OI but will be updated at every assimilation time step for EEWADAi and SUGOiWADAi. The correction to the model state $\mathbf{x}$ will be made using the observational data $\mathbf{y}$ (sampled from the control run) as

$$\mathbf{x} = \mathbf{x}^- + \mathbf{F}_{T \text{ or } S}\mathbf{K}[\mathbf{y} - \mathbf{H}\mathbf{x}^-], \qquad (27)$$

where $\mathbf{y}$ is always the SSH anomaly and $\mathbf{F}_{T \text{ or } S}$ is the statistical correlation between SSH anomaly and the temperature or salinity anomalies.

We have, therefore, tested the impact of varying $\mathbf{D}$ to the assimilation skill with the twin experiments. Two methods were implemented varying $\mathbf{D}$ based on the information obtained through the wavelet decomposition of the model output field. In practice, we have chosen the diag($\mathbf{D}$) as a function of the variation at three scales, see (22), and for each scale the variation was detected from the subspaces, $\mathbf{V}_j \otimes \mathbf{W}_j$ and $\mathbf{W}_j \otimes \mathbf{V}_j$, instead of the variation detected from the subspace $\mathbf{W}_j \otimes \mathbf{W}_j$ alone. In the former approach we can capture either horizontal edges or vertical edges, a less strict requirement for error detection than the latter. The weighting of the three scales was given as 1, 1/2, and 1/4 for SUGOiWADAi and 1, 0, and 0 for EEWADAi in ascending order in scale.

e. Results of the twin experiments

In the previous study (Jameson and Waseda 2000) we have shown that the new wavelet-based method (EE-WADAi) outperforms the traditional OI scheme after 120 days of assimilation. In other words the 2 solutions

were not noticeably different in quality up to day 120. Together with the previous results of EEWADAi, we present the L_2 difference of 2 of the flow fields, sea surface height and temperature at a given level, between the control run and the test run at 5-day intervals during the experiment of SUGOiWADAi. We can see from the two plots presented, see Figs. 6 and 7, that SUGOi-WADAi (red circles) reduced the error significantly around day 40–100 when EEWADAi (blue circles) gave a similar performance to OI (green circles). We also see that after 120 days both wavelet-based methods, EE-WADAi and SUGOiWADAi, have reached a kind of steady state.

It is clear now that this second-generation, wavelet-based assimilation scheme SUGOiWADAi has improved considerably over the earlier version EEWADAi. One can see from Figs. 6 and 7 that the red circles always show smaller errors than the others (blue and green circles) indicating the stability of the new skill. This probably is because the errors in the entire model domain are heterogeneous such that some part of the domain may have more numerical errors and some other part may have more errors due to model variation. The new method, SUGOiWADAi, will automatically detect both errors and weigh appropriately so that the impact of both can be incorporated into the estimation of the error covariance matrix at each time step. The success of this new scheme is rather important since for most reduced Kalman filtering approaches, the numerical errors and the errors associated with a localized model variation are neglected in order to save the computational load. We have shown in this study that such knowledge is essential in the improvement of the assimilation schemes.

5. Conclusions

In the field of data assimilation, one can summarize the weakness of current methods as saying there is in-

sufficient knowledge of errors, either from the computational side or from the external source of data. Without knowledge of errors one cannot hope to assimilate external data efficiently. From the simple example of incorporating real-time sea surface height data into a model run, one must know roughly which is more accurate, the height given by the numerical calculation or the height given by the external data source, say data from a satellite. Generally we will have some knowledge of the satellite errors but not as a function of space and time. For example, we might know that the satellite data error is some kind of skewed Gaussian with a certain standard deviation. But, we will generally have no knowledge of how this error changes with spatial location over the ocean or with time. On the other hand, if one has some knowledge of the errors in the computational scheme then one will have an estimate of the relative errors between the satellite data and computed data.

In this manuscript we have introduced a second-generation data assimilation scheme that uses wavelet analysis to build the error covariance for reduced Kalman filtering. In our first-generation scheme, EEWADAi, we used wavelets to detect numerical error. In our current second-generation scheme, SUgOiWADAi, we use wavelet analysis to not only detect numerical error but also to give estimates of local model variation at various scales. This utilization of information at various scales incorporates the strong points of EEWADAi with the strong points of the existing and commonly used optimal interpolation schemes. By bringing together the strong features of both approaches, we have created a very robust data assimilation method that outperforms both of the previous approaches. Furthermore, our new approach is computationally very inexpensive.

Our future intentions are to explore further ways to enhance SUgOiWADAi and to explore ways to broaden its application.

REFERENCES

Blumberg, A. F., and G. L. Mellor, 1983: Diagnostic and prognostic numerical circulation studies of the South Atlantic Bight. *J. Geophys. Res.,* **88,** 4579–4592.

Daubechies, I., 1988: Orthonormal basis of compactly supported wavelets. *Commun. Pure Appl. Math.,* **41,** 909–996.

Erlebacher, G., M. Y. Hussaini, and L. Jameson, Eds., 1996: *Wavelets: Theory and Applications.* Oxford University Press, 510 pp.

Ezer, T., and G. L. Mellor, 1994: Continuous assimilation of Geosat altimeter data into a three-dimensional primitive equation Gulf Stream model. *J. Phys. Oceanogr.,* **24,** 832–847.

Ghil, M., 1989: Meteorological data assimilation for oceanographers. Part I: Description and theoretical framework. *Dyn. Atmos. Oceans,* **13,** 171–218.

Jameson, L., 1998: A wavelet-optimized, very high order adaptive grid and order numerical method. *SIAM J. Sci. Comput.,* **19,** 1980–2013.

——, and T. Waseda, 2000: Error estimation using wavelet analysis for data assimilation: EEWADAi. *J. Atmos. Oceanic Technol.,* **17,** 1235–1246.

Kalman, R. E., 1960: A new approach to linear filtering and prediction problems. *J. Basic Eng.,* **82D,** 35–45.

Malanotte-Rizzoli, P., and W. R. Holland, 1986: Data constraints applied to models of the ocean general circulation. Part I: The steady case. *J. Phys. Oceanogr.,* **16,** 1665–1687.

Mellor, G., and T. Ezer, 1991: A Gulf Stream model and an altimetry assimilation scheme. *J. Geophys. Res.,* **96** (C5), 8779–8795.

Mitsudera, H., Y. Yoshikawa, B. Taguchi, and H. Nakamura, 1997: High-resolution Kuroshio/Oyashio System model: Preliminary results (in Japanese with English abstract). Japan Marine Science and Technology Center Rep. 36, 147–155.

Pham, D. T., J. Verron, and M. C. Roubaud, 1998: A singular evolutive extended Kalman filter for data assimilation in oceanography. *J. Mar. Syst.,* **16,** 323–340.