Wavelet-Compressed Representation of Deep Moist Convection

JUN-ICHI YANO*

CNRM, Météo-France, and CNRS, Toulouse, France

PETER BECHTOLD

ECMWF, Shinfield Park, Reading, United Kingdom

JEAN-LUC REDELSPERGER AND FRANCOISE GUICHARD

CNRM, Météo-France, and CNRS, Toulouse, France

(Manuscript received 7 February 2003, in final form 15 December 2003)

ABSTRACT

The capacity of wavelets to effectively represent atmospheric processes under compression is tested by a dataset generated by a cloud-resolving model simulation of deep convective events observed during the Tropical Ocean Global Atmosphere Coupled Ocean–Atmosphere Response Experiment (TOGA COARE).

Generally, more than 90% of the total variance is reproduced by retaining only the top 10% of the modes. The compressed data does not drastically deteriorate for graphic purposes, even when only the top 1% of modes are retained. The performance of compression is overall comparable for all the wavelets considered and also does not strongly depend on the type of physical variables.

Conventional quantitative measures do not distinguish the compression errors arising from different characters of individual wavelets well, although different wavelet modes filter out different structures as ''noises'' depending on their characteristics. The importance of choosing wavelets that represent the shape of signals physically expected is emphasized. Analytical discontinuities of wavelets are not necessarily undesirable for all the purposes, but must be consistent with our physical picture for the system. For this reason, the Haar wavelet may be acceptable because of its piecewise-constant structure, whereas the Daubechies wavelets of lower degrees are less appropriate because of their highly irregular structures.

Some preliminary analyses are performed for assessing the capacity of wavelets to represent the full physics of meteorological systems. It is suggested that the loss of magnitudes in vertical fluxes under high compression can be recovered by a kind of ''renormalization'' to a good extent. The mass continuity is found to be reasonably satisfied under the proposed compression method, although the latter is not explicitly constrained by the former.

1. Introduction

With our increase both in computing power and observational capabilities, the amount of available meteorological data also increases dramatically. Efficient handling of these datasets, including their effective compression and representations, is becoming more and more important.

For example, in high-resolution numerical modeling, the accumulating output is so large that it is usually stored only with a relatively long time increment. Consequently, it is very often hard to access the details of the temporal evolution of a simulation, in spite of their potential importance. A similar problem is posed by satellite data: with their increasing spatial resolution (a notable example is given by geostationary satellite images), it is increasingly hard to utilize them in their full resolutions both in space and time simultaneously.

Data compression by wavelets is one possibility to be considered in response to these needs. Wavelets may already be considered as an established standard method in meteorology. The wavelet analysis has been applied in the tropical meteorology including ENSO (Weng and Lau 1994; Kestin et al. 1998), ITCZ (Gu and Zhang 2001), intraseasonal variability (Tung and Yanai 2002), and African easterly waves (Diedhiou et al. 1999; Grist 2002). Other meteorological applications include, for example, stratospheric gravity waves (Sato and Yamada 1994), blocking (Fournier 2002, 2003), and the boundary layer (Galmarini and Attié 2000; Attié and Durand 2003). Various introductory reviews for meteorologists

* Current affiliation: Laboratoire de Météorologie Dynamique, Université Pierre et Marie Curie, Paris, France.

Corresponding author address: Jun-Ichi Yano, CNRM, Météo-France, 42 av. Coriolis, 31057 Toulouse Cedex, France.
E-mail: yano@cnrm.meteo.fr

have been provided by Mak (1995), Lau and Weng (1995), Torrence and Compo (1998), and Fournier (2000); the reader is referred to Mallat (1998) as a general reference for the subject.

However, the wavelets' capacity for efficiently compressing the data is not yet widely recognized in our community [but see Marcal et al. (2000) as an exception], although several convincing demonstrative examples and quantitative comparisons are provided by Fournier (2000, see his Figs. 1, 7, 8, and 9), and this possibility is quantitatively supported by the probability distribution of wavelet coefficients by Yano et al. (2001b).

Issues of data compression can be further divided into ''compression'' in the narrow sense and wider ''representation'' issues. The former simply refers to a technique of storage of the essential content, in some sense, of a dataset. Its chief goal is to reduce the size of the dataset without losing that essential content. Moving averages in space and time, or equivalently, filtering in the Fourier space is a traditional way to achieve this goal by simply neglecting fine structures of the original data. The wavelet can compress data more effectively without substantially losing fine structures.

In contrast, ''representation'' refers to the problem of efficiently representing the physical system by a set of basis functions. For example, the Fourier and the spherical harmonics are widely used in global numerical models and the storage of the global analysis and reanalysis datasets. But the standard mode expansion method does not work as a compression, because generally the data information is lost with increasing truncation rate. The best-known example of the mode expansion method in our community that also works as a compression is the empirical orthogonal functions (EOFs) whose few leading modes can often explain a large part (say more than 90%) of the climate variability. However, because of a fixed spatial pattern used in the representation, fine structures of the variability are inevitably lost by this method.

In the present paper, the discrete orthogonal wavelet transform is suggested for this purpose, which can be directly used in place of the Fourier modes in representing physical systems (cf. Vasilyev et al. 1997). The time-dependent ''sparse'' truncation of these discrete wavelet modes can more effectively represent the meteorological system. On the other hand, the continuous wavelets are constructed in a highly redundant manner and, although useful for statistical analyses, are not convenient for such representations [see, e.g., Fournier (2000) for more extensive discussions].

The ''representation'' capability of discrete wavelets leads to the possibility of formulating a meteorological system in terms of a wavelet expansion, in place of, say, the Fourier or spherical harmonics, for various diagnoses such as energy cycle analyses (Fournier 2002, 2003), as well as for prognostic studies (Ekstrom and Hales 2000). By extending these studies, the present authors would like to construct a prognostic model for moist convection based on a wavelet-compressed representation. A highly compressed version of this model would provide subgrid-scale representations of deep convection for global models based on its compactness. This approach follows a similar line of thought as the superparameterization proposed by Randall et al. (2003), but the resulting model would be of much lower dimensions.

With this final goal in mind, we investigate the compressibility by wavelets of a tropical deep convective system simulated by a cloud-resolving model (CRM). The present work extends earlier work by Yano et al. (2001a,b) that demonstrated the usefulness of the wavelets in analyzing CRM simulation data. The Meyer wavelet was chosen in this earlier work. Here, in addition to the Meyer wavelet, various wavelets under the general framework of the Mallat algorithm (cf. Mallat 1998, chapter 7) are applied in order to better explore the possibilities of wavelet applications to CRM.

The paper is organized as follows: The wavelet method is outlined in section 2 with general applications in mind. It is followed by a brief presentation of the CRM simulations in section 3. Compression tests on the simulated fields including vertical fluxes are discussed in sections 4 and 5, and the mass conservation issue is addressed in section 6. Finally, the conclusions from these analyses and practical applications are summarized in section 7.

2. Wavelet method

a. Discrete orthogonal wavelets

Discrete orthogonal wavelets constitute a set of orthogonal complete functions that can be used to represent any well-behaved function $\varphi(x)$ in a periodic domain $[0, L]$.[1] They are constructed in a manner to provide characterizations for both the scale and the localizations of a field. These two characteristics are designated by the two indices, i and j, respectively. Thus, when N equidistant data are given, the space-scale index spans for $i = 0, 1, 2, \ldots, i_{\max} \equiv \log_2 N - 1$ corresponding to the scale $L/2^i$, *or* the wavenumber band $k = 2^i$. Here, the maximum wavenumber band is $k_{\max} = N/2$. For a given wavenumber band k, there are k linearly independent wavelets $\psi_l(x)$ that are constructed by shifting their peaks. These are designated by $j = 0, \ldots, k - 1$. Here, the single index l for characterizing wavelet modes is introduced by

$$l = k + j. \tag{2.1}$$

Finally, in order to make this set complete, the mean component is defined by

$$\psi_0(x) = 1,$$

[1] For the treatments of the nonperiodic boundary conditions, we refer to Daubechies (1994) and Mallat (1998, section 7.5).

where one takes $k = 0$ (and $j = 0$) for convenience. As a result, the index l spans from 0 to $N - 1$. Note that the domain mean vanishes for the remaining wavelet modes.[2]

Thus, a function $\varphi(x)$ can be represented in terms of wavelets as

$$\varphi(x) = \sum_{l=0}^{N-1} \hat{\varphi}_l \psi_l(x). \tag{2.2}$$

By using the orthogonality

$$\frac{1}{L}\int_0^L \psi_l(x)\psi_{l'}(x)\, dx = \delta_{ll'} \tag{2.3}$$

the expansion coefficients are computed by

$$\hat{\varphi}_l = \frac{1}{L}\int_0^L \varphi(x)\psi_l(x)\, dx, \tag{2.4}$$

with δ_{ij} the Kronecker's symbol.

As the present CRM has doubly periodic Cartesian coordinates (x, y) in the horizontal domain, the wavelet expansion is given by

$$\varphi(x, y) = \sum_{l_x=0}^{N_x-1}\sum_{l_y=0}^{N_y-1} \hat{\varphi}_{l_x,l_y}\psi_{l_x}(x)\psi_{l_y}(y), \tag{2.5}$$

where N_x and N_y are the numbers of data points in the x and y directions, respectively. The expansion coefficients $\hat{\varphi}_{l_x,l_y}$ are given analogously to Eq. (2.4).

Note that deep moist convection simulation data consists of three spatial dimensions, but the wavelet decomposition is applied only on the horizontal directions. Although the standard meteorological data is given in the three spatial dimensions, in most cases it is found to be less convenient to expand the variables by normal modes (cf. Kasahara and Puri 1981; Fulton and Schubert 1985) in the vertical direction z.

The above formulation can in principle also be applied to global datasets defined over a sphere, if care is taken on the change of metrics in the differential operators. Alternatively, the wavelets may be redefined in the spherical coordinates, which are currently under active investigations (e.g., Schröder and Sweldens 1995; Li 1999).

b. Choice of wavelet types

The mode decomposition constitutes a pattern-fitting problem. Therefore, the shape of the wavelets must represent the character of the expected physical signals (cf. Torrence and Compo 1998; see also Qiu et al. 1995; Desrochers and Yee 1999).

The three types of wavelets have been selected: the Meyer wavelet, the Haar wavelet, and the Daubechies wavelet of degree $p = 2$. They are illustrated in Fig. 1 for the wavenumber band $k = 4$. The Meyer wavelet takes the shape of isolated wave packets with an analytical continuity (i.e., no discontinuities in derivatives) that are expected to well represent pulselike signals. The Haar wavelet consists of a pair of square pulses, which may be considered as a piecewise-constant idealization of pairs of ''downdraft'' and ''updraft.'' The Daubechies wavelet takes a highly irregular shape, which may be relevant in representing highly intermittent turbulence signals.

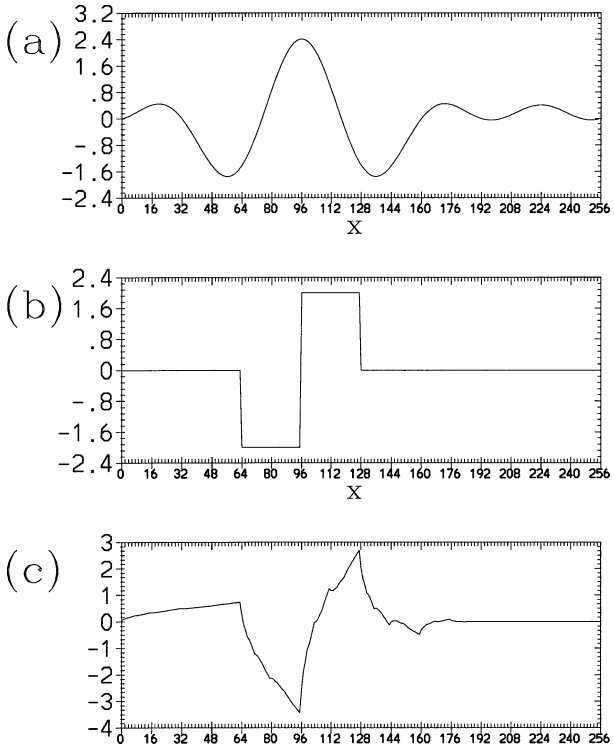

FIG. 1. Samples of wavelets with wavenumber band $k = 4$ as a function of the horizontal coordinate: (a) Meyer, (b) Haar, and (c) Daubechies wavelets of degree $p = 2$.

The Meyer wavelet is computed in Fourier space as detailed in Yamada and Ohkitani (1991). The Haar and Daubechies wavelets are computed by the Mallat algorithm [Mallat 1998, chapter 7; see also a summary in Fournier 2000, section 2a; Jameson and Waseda 2000, section 2; Jameson et al. 2002, section 2; Luo and Jameson 2002, section 2]. Other types of wavelets (Daubechies of different degrees as well as Battle-Lemarié spline wavelets, Coiflets) are also examined under the Mallat algorithm. We refer to Mallat (1998, chapter 7) for mathematical definitions and characters of these wavelets, in general, and more specifically on computational efficiency (see section 7.3.1 therein), regularity, and compactness in both real and Fourier spaces (see sections 7.2 and 7.3 therein).

[2] More precisely, this very last mode corresponds to a leading mode ϕ_0 for scaling functions (cf. Fournier 2000, section 2d; Mallat 1998, chapter 7).

c. Data compression

The basic idea of data compression in wavelet space is based on the observation that the accuracy of the expansion (2.2) may only be weakly affected by removing the small coefficient $\hat{\varphi}_l$, when the variable φ is smooth enough. As a result, only n sparsely distributed modes defined by the indices $l(q)$ with $q = 1, \ldots, n$ are retained, and the expansion is approximately given by

$$\varphi(x) \simeq \sum_{q=1}^{n} \hat{\varphi}_{l(q)} \psi_{l(q)}(x). \tag{2.6}$$

The ratio n/N is referred to as the compression rate.

Wavelet data compression is a rapidly developing field [see Mallat (1998, chapter 9) for an overview; Chang et al. (2002) as an example of more recent developments; and Marcal et al. (2000) for a good introductory review from a perspective of image compression]. Here, the simplest available compression is used, in which only the modes with absolute values larger than a threshold $\hat{\varphi}_c$ are retained:

$$|\hat{\varphi}_l| > \hat{\varphi}_c. \tag{2.7a}$$

The threshold $\hat{\varphi}_c$ may be scale dependent in general [see Yano et al. (2004) for such an example], but we simply set it proportional to the standard deviation $\langle \varphi'^2 \rangle^{1/2} \equiv (\sum_{l=1}^{N-1} \hat{\varphi}_l^2)^{1/2}$ [i.e., Plancherel formula; cf. theorem 2.3 of Mallat (1998)] of φ:

$$\hat{\varphi}_c = \mu(\langle \varphi'^2 \rangle / N)^{1/2}, \tag{2.7b}$$

with μ a constant factor to specify the degree of compression.

The same principle also applies to two-dimensional datasets. Hence, the compressed two-dimensional data is given by

$$\varphi(x, y) \simeq \sum_{q=1}^{n} \hat{\varphi}_{l_x(q), l_y(q)} \psi_{l_x(q)}(x) \psi_{l_y(q)}(y), \tag{2.8}$$

where $[l_x(q), l_y(q)]$ $(q = 1, \ldots, n)$ designates the indices for the retained modes.

For $\mu = 1$ the thresholding method reduces to the Lorenz threshold considered by Katul and Vidakovic (1998). This provides an optimized threshold in the sense that beyond this threshold, the rate of loss of variance by truncation becomes faster than the rate of decrease in compression rate. An alternative thresholding method based on an estimate of the variance of the small-scale noises has been proposed by Donoho and Johnstone (1994) and applied to turbulence simulations by Farge et al. (1999).

3. Outline of simulation

The CRM simulations were performed with the nonhydrostatic mesoscale model (Meso-NH) developed jointly by the Laboratoire d'Aerologie and Centre National de Recherches Meteorologiques/Groupe d'étude de l'Atmosphère Météorologique (CNRM/GAME) (cf. Lafore et al. 1998). The simulation covers a 2-day period from 1200 UTC 10 December to 1200 UTC 12 December 1992 during the Tropical Ocean Global Atmosphere Coupled Ocean–Atmosphere Response Experiment (TOGA COARE; see Guichard et al. 2000). The model extends over a relatively large domain of 256 × 256 equidistant grid points with a spacing of 2 km; the boundary conditions are biperiodic. The model is forced by observed ''large-scale'' temperature and moisture tendencies, and the winds are relaxed toward the domain-mean observed values (cf. Grabowski et al. 1996). For the present analysis, only tropospheric data from the 42 model levels below 18 km is used. The vertical levels have finer resolutions close to the surface (70 m at the lowest) and are gradually stretched to the crudest spacing of 0.7 km at the upper troposphere. The model runs on a f plane assuming the latitude 2.0°S.

The analyses are considered for the three components of the wind vector (u, v, and w) (labeled hereafter 1, 2, and 3), the potential temperature θ (4), the water vapor mixing ration q (5), and the total condensate mixing ratio q_c (6). The original data are given on a staggered grid (C grids; Mesinger and Arakawa 1976) in the three-dimensional space. In order to evaluate the vertical momentum fluxes in the wavelet space in section 4, the horizontal wind velocities have been interpolated to the same horizontal grid as for the thermodynamic variables. Additionally, the vertical component of the vorticity ζ (variable 0) is computed using the horizontal winds on the original staggered grid.

A snapshot after 30 h of the integration, corresponding to the maximum domain-averaged precipitation rate, is used as an illustrative example. Figures 2 and 3 show respectively a horizontal and vertical cross section of the total condensate and wind fields. The deep convection appears to be loosely organized in some larger clusters. [This snapshot is used later in Figs. 4–9, and 13–15, whereas for some statistics (Figs. 10–12) time averaging is performed over 16 samples of the whole experiment period (48 h with a 3-hourly sampling)].

4. Compressibility tests

The wavelet compression described in section 2c is applied to the CRM data separately for each vertical level. The standard deviation $\langle \varphi'^2 \rangle^{1/2}(z) \equiv [\sum_{l_x=0}^{N_x-1} \sum_{l_y=0}^{N_y-1} \hat{\varphi}'^2_{l_x,l_y}(z)]^{1/2}$ is also evaluated at each vertical level, whereas the threshold factor μ is kept constant with height. Here, $N_x = N_y = 256$ are the number of grid points in x and y directions. The compression rate is n/N, with n the number of retained modes, and $N = N_x N_y = 65\,536$.

a. Examples

The degree of compression is controlled by the threshold factor μ introduced in Eq. (2.7b). Compression rates

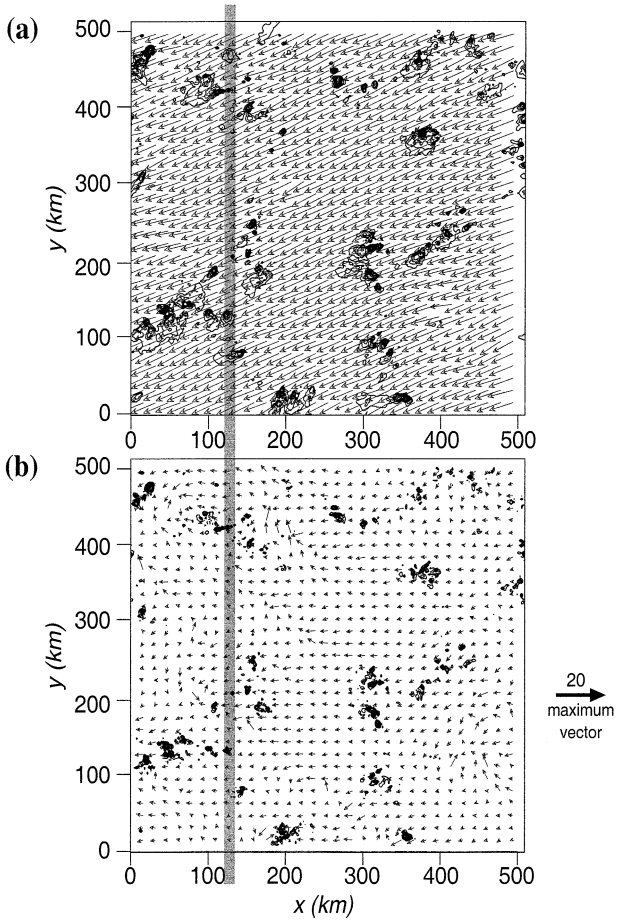

FIG. 2. Top view of the convective simulation at $t = 30$ h: (a) total condensate and (b) vertical velocity. Both variables are arithmetically averaged over all the model vertical levels (thus weighted by coordinate stretching) and shown with the contour intervals of 0.5 g kg^{-1} and 0.5 m s^{-1}, respectively. Also shown are the vertically averaged horizontal wind vectors (with the unit length for 20 m s^{-1} shown in the lower right) for the upper vertical levels (above 5 km) and the lower vertical levels (below 5 km) in (a) and (b), respectively. Shading indicates location of vertical cross section in Figs. 3–9.

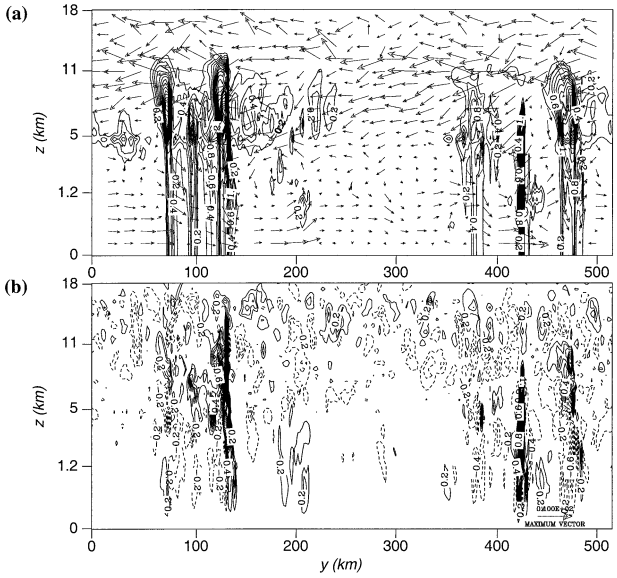

FIG. 3. Vertical cross section averaged over $x = 120$–130 km (see the shading in Fig. 2) of (a) the total condensate and (b) the vertical velocity, with contour intervals of 0.2 g kg^{-1} and 0.2 m s^{-1}, respectively; (v, w) wind vectors are also shown in (a) (unit length 10 m s^{-1}, with the vertical wind multiplied by 10). The vertical coordinate is stretched following the model vertical levels, with the middle level corresponding to 5-km height.

on the order of 10% and 1% have been obtained with a threshold factor $\mu = 1$ and $\mu = 5$, respectively. Examples of compression under the threshold factor $\mu = 5$ are shown in Figs. 4–6 for the wavelets designated Meyer, Haar, and Daubechies of degree $p = 2$, respectively. The same vertical cross sections as in Fig. 3 are used. Even under this large compression rate of 1% all wavelets retain the main convective structures and circulations. Subjectively, the degradation of the fields seems to depend on the physical variables as well as wavelet types, with the vertical velocity field being most affected. Especially, the vertical extension of convective updrafts in the cross-section plots is substantially shortened by heavy compression, but less so by Daubechies. The total condensate field is reproduced by the Haar wavelet with highly homogenized features, reflecting that wavelet character, whereas a smooth field is produced by the Meyer wavelet.

These features are identified more clearly in Figs. 7–9, where the total condensate mixing ratio q_c and the vertical velocity w are plotted at $z = 3.8$ km along the

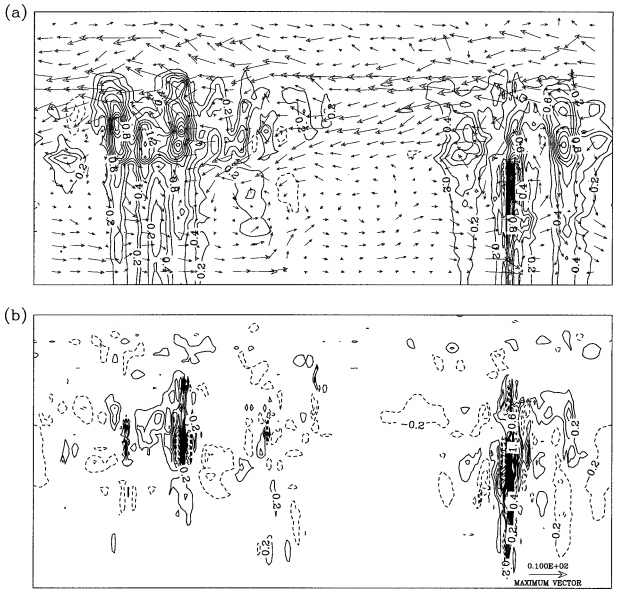

FIG. 4. The same as Fig. 3, but under the compression by the Meyer wavelet with the compression factor $\mu = 5$. Less than 8×10^{-3} of the modes are retained here and in the followings plots (Figs. 5–9; cf. Figs. 10a and 11a).

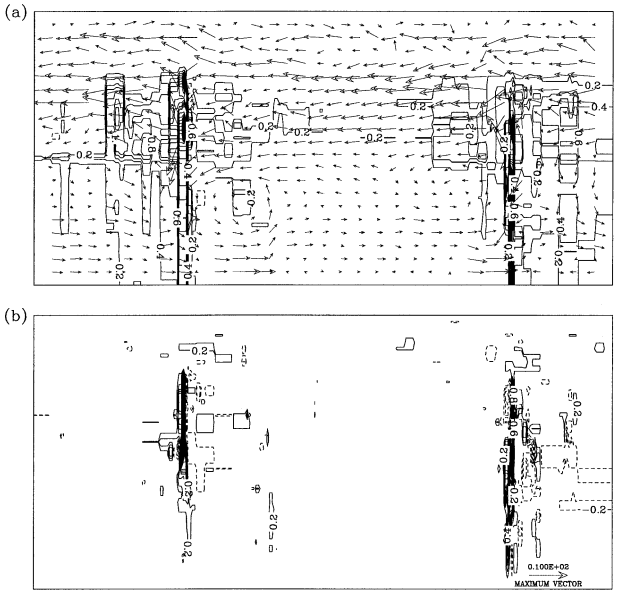

FIG. 5. The same as Fig. 4, but with the Haar wavelet.

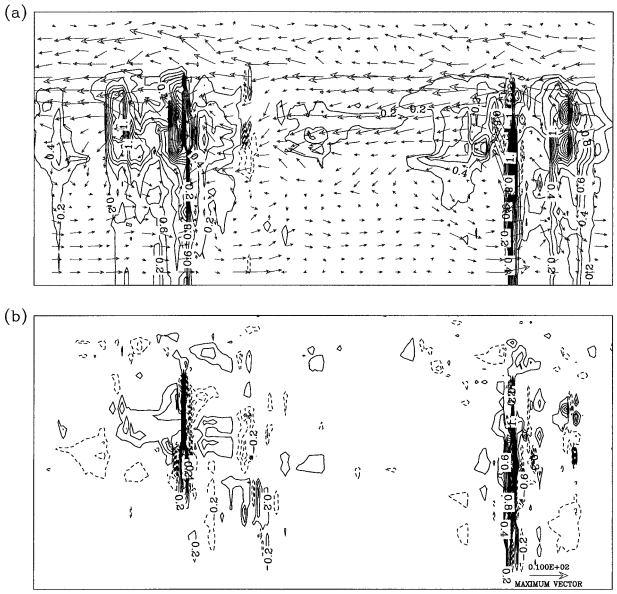

FIG. 6. The same as Fig. 4, but with the Daubechies wavelet of degree $p = 2$.

y section at $x = 120$ km. In spite of an obvious degradation of the original data, the wavelet-compressed data clearly reproduces the overall features of the original physical fields. Especially, a strong updraft feature at $y \simeq 420$ km associated with a high condensate content is well retained by the high compression with $\mu = 5$ for all wavelets. However, a major concern in compressing the physical fields by wavelets also becomes clear. The positive-definite variables such as the total condensate may exhibit negative values after compression. The Meyer wavelet, because of its tendency to produce long wavy tails, produces several negative bumps in the zero–q_c range ($y = 120$–330 km in Fig. 7). This undesirable feature is further enhanced with the Daubechies wavelet of order $p = 2$, with negative extrema more pronounced, but it reproduces better the maximum values (Fig. 9). The Haar wavelet generates the least artificial edge-wave-like features and reproduces the zero–q_c region with the least bumps (Fig. 8). This aspect makes the Haar wavelet more favorable than the other wavelets.

b. *Quantifications*

As an attempt to quantify the compression capabilities, the explained variance, the rms error, and the maximum error (the maximum absolute difference from the original data) are plotted in Figs. 10 and 11 as a function of the compression rate. The relation between the compression rate and the threshold factor μ is also depicted in Figs. 10a and 11a. All the statistics are given in terms of the averages over all vertical levels and all 16 3-hourly time snapshots. The explained variance is given by the ratio of the vertically averaged variance of the compressed data to that of the original, as defined for

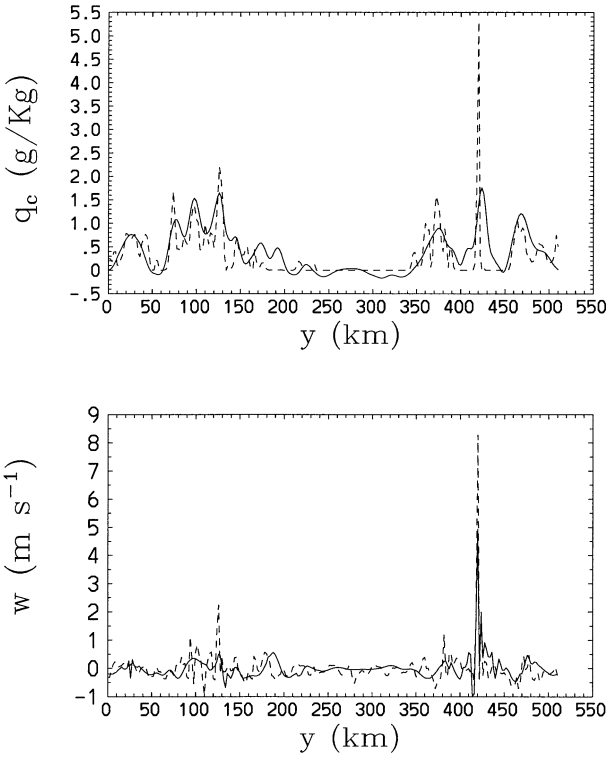

FIG. 7. Distributions of the (a) condensate mixing ratio q_c and the (b) vertical velocity w at $z = 3.8$ km along the y section at $x = 120$ km. The original data (short dash) is shown along with the data (solid) compressed by the Meyer wavelet with the compression factor $\mu = 5$, corresponding to a compression rate $n/N < 8 \times 10^{-3}$.

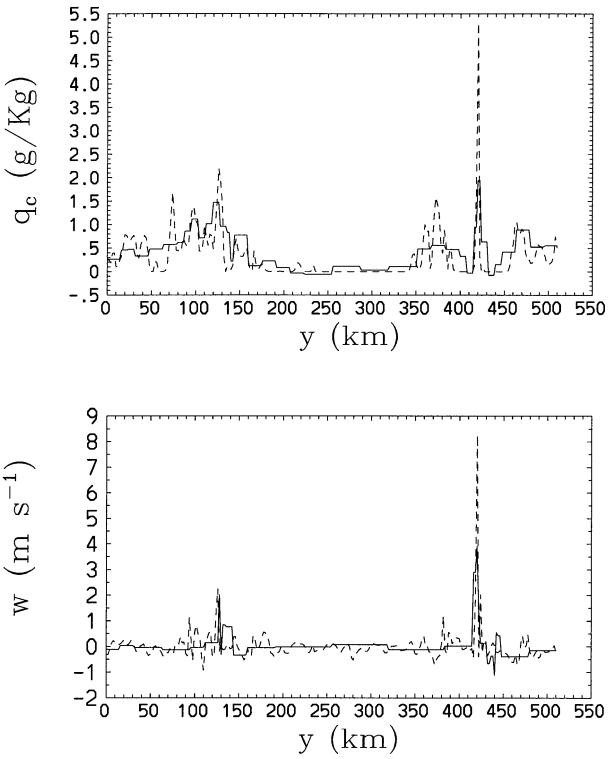

FIG. 8. The same as Fig. 7, but with the Haar wavelet.

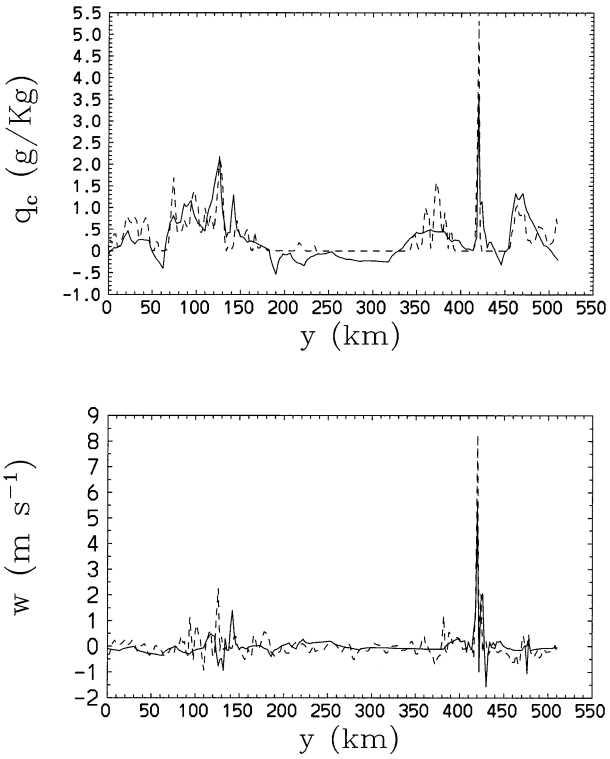

FIG. 9. The same as Fig. 7, but with the Daubechies wavelet of degree $p = 2$.

each snapshot. The maximum error refers to the highest error found in the three-dimensional data. Note that the variance is computed after subtracting the horizontal mean at each vertical level. Errors (rms and maximum) are normalized by the standard deviation over the three-dimensional domain (defined as a deviation from a horizontal mean at each vertical level). For reference, the standard deviations for the analyzed seven variables at $t = 30$ h are listed in Table 1. The variability of the potential temperature is rather weak, as is typically the case for the tropical atmosphere (cf. Charney 1963).

Figure 10 shows the results for the Meyer wavelet, where the curve numbers correspond to the variable numbers. The compressibility clearly depends on the physical variable; explained variance decays the slowest with decreasing n/N for the potential temperature (4) and the horizontal winds (1, 2), maintaining the explained variance above 95% and 75%, respectively, down to the 1% compression. This reflects the fact that these variables are dominated by the large-scale field. For such a field, the total variance does not reduce substantially by removing the small-scale substructures. The vorticity (0) is most affected by the compression, as it is dominated by the small-scale features. The compressibility of the velocity (3), moisture (5), and total condensate (6) fields lies in between the above two cases. These fields are dominated by the localized cloud–convection-related features for which the wavelet is expected to be well adapted. For these variables, more than 90% of the total variance is maintained under 10% compression, and more than 50% under 1% compression. The overall compressibility of the deep convective wind and temperature data is similar to that found for atmospheric boundary layer data by Katul and Vidakovic (1998), although the compressibility appears to be slightly lower in the present case.

We have also compared the compressibility of selected physical variables by the various wavelets. The data compressibility is surprisingly similar for all the wavelets tested, in spite of their different characters. Such an example is shown for the vertical velocity in Fig. 11. Superiority or inferiority of a particular wavelet appearing in individual variables is not consistent and is likely to be insignificant. For example, the Haar wavelet (curve 2) behaves worst for the zonal wind and moisture (not shown), but it appears to perform better along with the Daubechies ($p = 2$; curve 3) for the vertical velocity. Most importantly, unfavorable features seen in the direct plots of the physical quantity (Figs. 3–9) do not clearly reflect on the compressibility as measured, for example, by the explained variance.

These compressibility tests have also been performed for some selective snapshots from this TOGA COARE CRM simulation, as well as with CRM data of a simulation of continental convection observed over the central United States during the Atmospheric Radiation Measurement (ARM) experiment (Xie et al. 2002). The

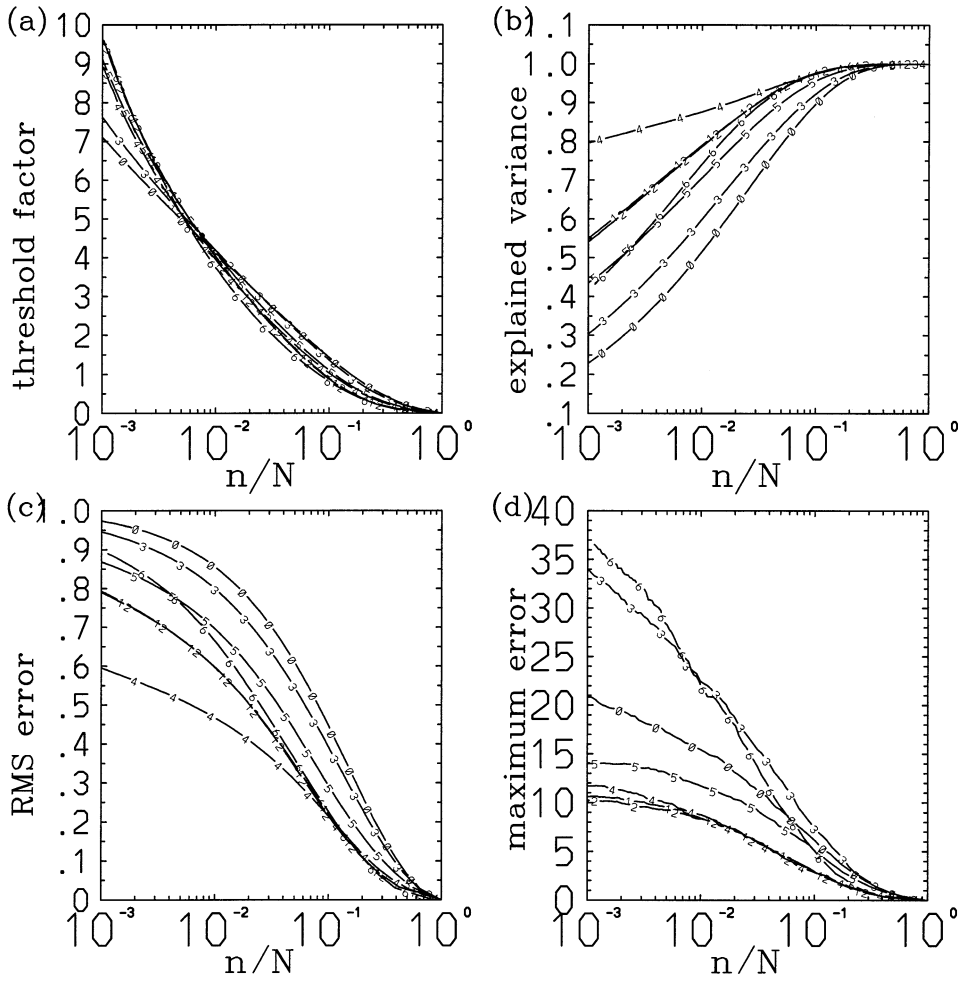

FIG. 10. Compression tests with Meyer wavelets for the TOGA COARE CRM simulation. Statistics are averaged over the 48-h simulation with 3-hourly data sample and represented as functions of the compression rate n/N: (a) the compression factor μ, (b) the explained variance, (c) the rms normalized error, and (d) the normalized maximum error. The curve numbers correspond to the vertical vorticity ζ (0), the three components of the winds (u, v, and w) (variables 1, 2, and 3), potential temperature θ (4), moisture mixing ratio q (5), and the total condensate mixing ratio q_c (6).

compression rate of data by wavelets was found to be not sensitive to the phases of convection.

5. Vertical flux representation

The previous section has demonstrated the ability of wavelets to efficiently compress individual variables in convective systems. However, a wavelet-compressed system, as defined by Eq. (2.8), must also faithfully reproduce the various physical processes such as the vertical fluxes of momentum, heat, and moisture. These vertical transports within convective systems significantly influence the evolution of large-scale/synoptic-scale systems. In biperiodic CRM simulations these processes are typically defined as domain-mean values, whereas in global circulation models these processes are represented by subgrid parameterizations.

Using (2.3) and (2.5) the domain-averaged vertical flux of a variable φ is given by

$$\overline{w\varphi}(z) \equiv \frac{1}{L_x L_y}\int_0^{L_x}\int_0^{L_y} w\varphi \, dx \, dy = \sum_{l_x=0}^{N_x-1}\sum_{l_y=0}^{N_y-1} \hat{w}_{l_x,l_y}\hat{\varphi}_{l_x,l_y}. \quad (5.1)$$

The overbar designates the horizontal domain mean, and $\overline{w} = 0$ (thus $\hat{w}_{0,0} = 0$) in the present stimulation. Following the definition (2.8), the domain-averaged vertical flux is approximated as

$$\overline{w\varphi}(z) \simeq \overline{w\varphi}^{\mu}(z) \equiv \sum_{(l_x,l_y)\in \mathbf{Q}_w\cap\mathbf{Q}_\varphi} \hat{w}_{l_x,l_y}\hat{\varphi}_{l_x,l_y}, \quad (5.2)$$

with the superscript μ standing for the threshold factor used in the compression. Here, $\mathbf{Q}_w \equiv \{(l_x, l_y): |\hat{w}_{l_x,l_y}|$

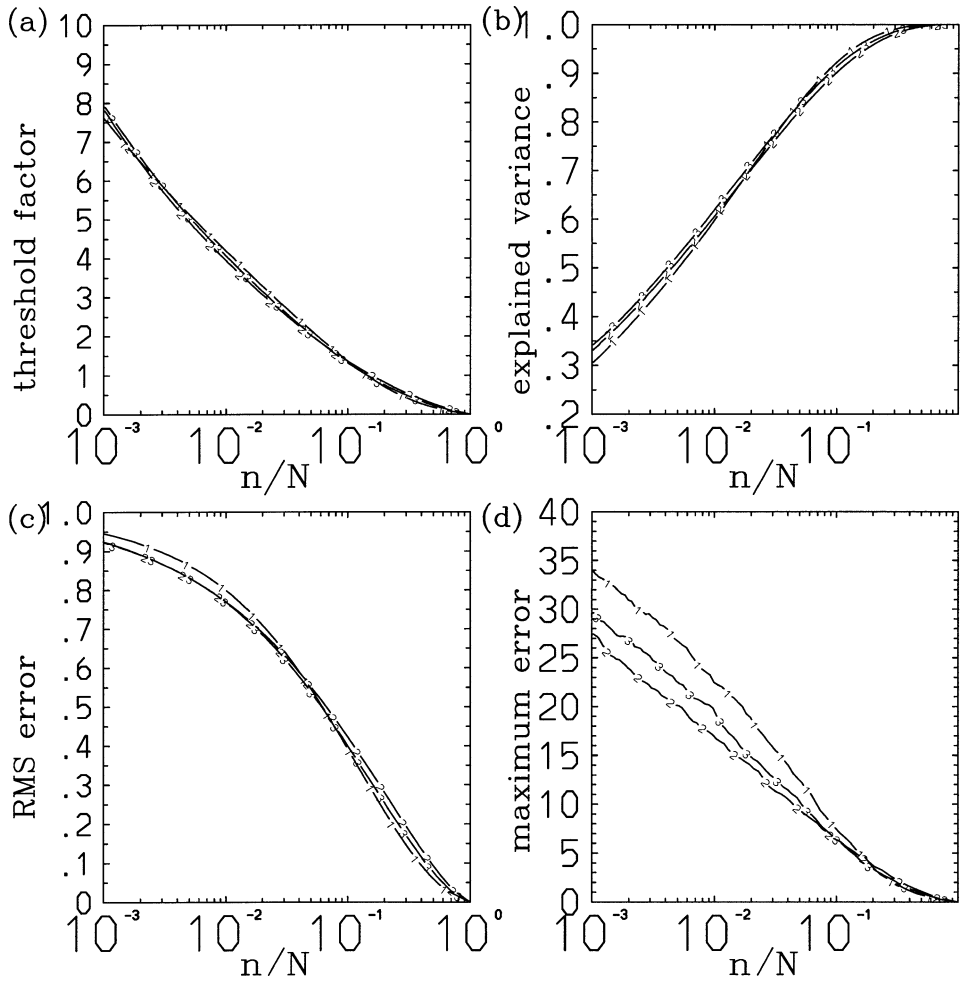

FIG. 11. The same as Fig. 10, but for the vertical velocity with the Meyer (1), the Haar (2), and the Daubechies wavelet with $p = 2$ (3).

$> \mu(\langle w'^2\rangle/N_xN_y)^{1/2}\}$, $\mathbf{Q}_\varphi \equiv \{(l_x, l_y): |\hat{\varphi}_{l_x,l_y}| > \mu(\langle w'^2\rangle/N_xN_y)^{1/2}\}$ are the sets of indices (l_x, l_y) retained by the compression for w and φ, respectively. Thus, the sum in Eq. (5.2) is taken over the intersection of these two sets (i.e., over the modes retained both for w and φ). On the other hand, in measuring the compression rate n/N in the following, we take a number of elements n in $\mathbf{Q}_w \cup \mathbf{Q}_\varphi$ (i.e., the number of modes retained either for w or φ) in order to indicate the total size of the truncated model system with these two variables.

TABLE 1. Standard deviations, defined as a deviation from a horizontal mean at each vertical level, of the analyzed variables at $t = 30$ h.

Variable no.	Variable	Std dev
0	ζ	6.4×10^{-4} s^{-1}
1	u	1.9 m s^{-1}
2	v	1.9 m s^{-1}
3	w	0.53 m s^{-1}
4	θ	1.7 K
5	q	0.60 g kg^{-1}
6	q_c	0.48 g kg^{-1}

The reproducibility of the vertical flux may be measured by

$$\alpha(\mu) \equiv \frac{\int_0^H \rho \ \mathrm{sgn}[\overline{w\varphi}(z)]\overline{w\varphi}^\mu(z)\, dz}{\int_0^H \rho|\overline{w\varphi}(z)|\, dz}, \qquad (5.3)$$

with $H = 18$ km the top of the analysis domain, and ρ the reference density. This is essentially the ratio of the vertically averaged compressed and absolute values. The factor $\mathrm{sgn}[(\overline{w\varphi}(z))]$ assures that this reproducibility measure counts as positive only if the compressed flux has the same sign as that of the original data.

This reproducibility measure is plotted in Fig. 12 for the vertical flux of the (a) and (b) u and v velocity

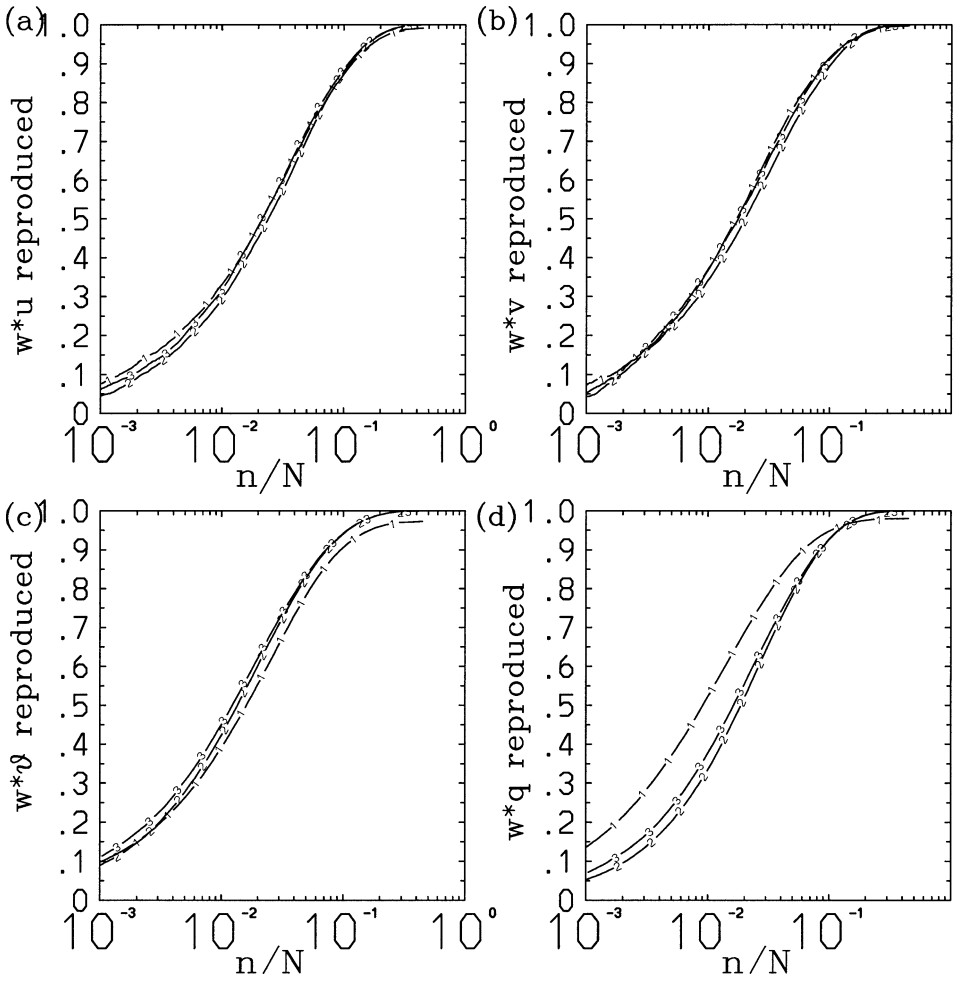

FIG. 12. Reproducibility of domain-mean vertical fluxes under compression by various wavelets as measured by Eq. (5.3), which is plotted against the compression rate n/N. (a) Zonal momentum flux, (b) meridional momentum flux, (c) heat flux, and (d) moisture flux. Wavelets considered are the Meyer (1), the Haar (2), and Daubechies with $p = 2$ (3).

components, respectively, (c) heat, and (d) moisture as functions of compression rate for selected wavelet types. The compression of different wavelets depends sensitively on the type of the vertical fluxes. The Meyer wavelet (curve 1) clearly outperforms the others for the moisture flux. On the other hand, both the Haar (curve 2) and the Daubechies with $p = 2$ (curve 3) perform slightly better than the Meyer wavelet for the vertical heat flux. Among the same class of wavelets (Daubechies and Coiflets), the more compact, less smooth version (of lower degree p) uniformally performs better for the heat and the moisture fluxes (not shown). But no monotonous dependence on the degree p is found for the momentum fluxes. Overall, there is no clear winner among the tested wavelets for the efficient representation of the vertical fluxes. Generally, more than 90% of the total flux is explained by retaining only 10% of the total modes, which constitutes a similar degree of compressibility as was the case for the individual variables.

The capacity of wavelet data compression is further investigated in Fig. 13, where the domain-mean vertical flux profiles are plotted for the Meyer wavelet as a function of the compression factor μ. The overall reduction of the vertical fluxes remains 10%–20% under a compression rate of 10% ($\mu = 1$: long dash; $\mu = 2$: short dash). The representation quality is clearly superior to that of the conventional mass-flux-based analysis (see Fig. 16 of Guichard et al. 1997). Nevertheless, a gradual reduction of the fluxes with increasing compressions is inevitable: with the 1% compression ($\mu = 5$), the amplitude of the flux typically reduces to less than one-half (chain dash). However, the shape of the flux profiles is essentially conserved under the compression. We attribute this tendency to the vertical coherence of convective systems. As a result, the wavelet compression loses variability in the vertical coherency with increas-

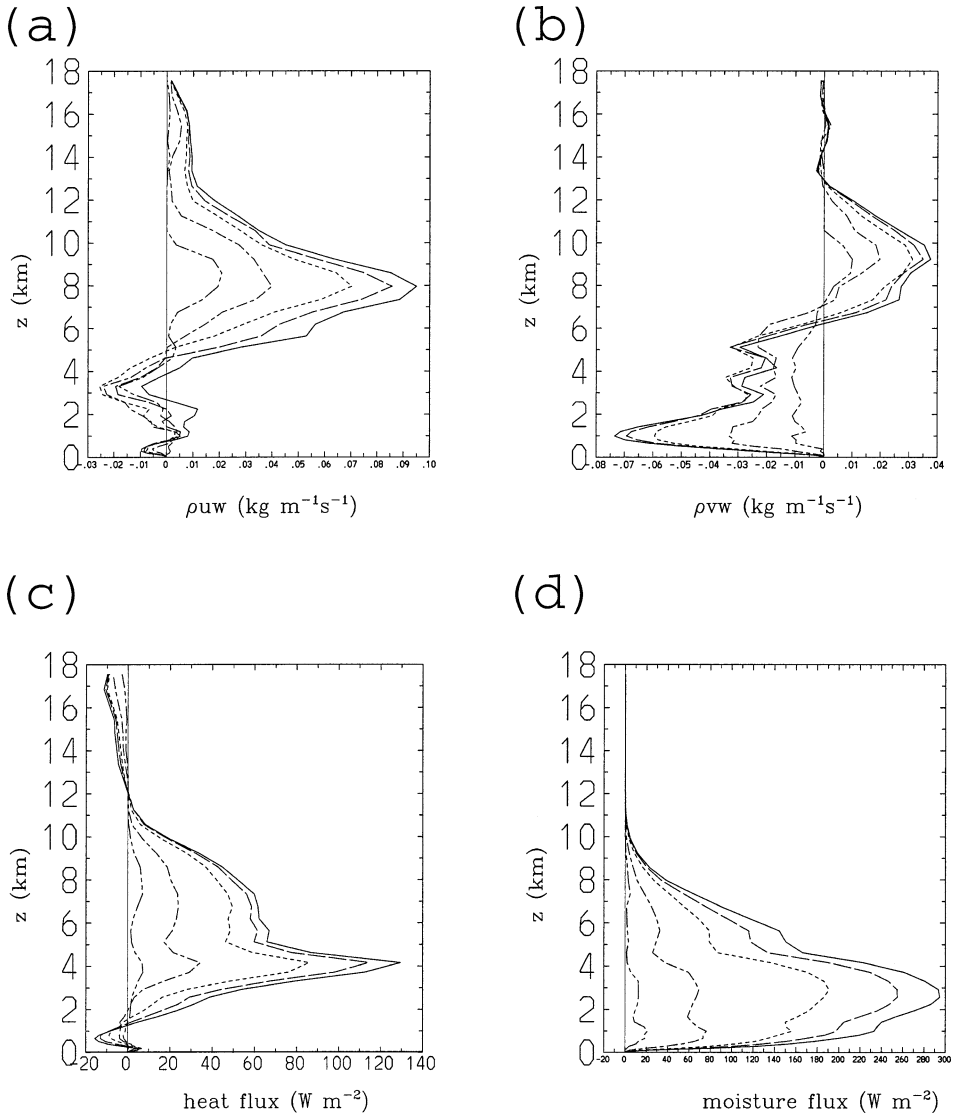

FIG. 13. Vertical flux profiles under the compression with the Meyer wavelet using a compression factor $\mu = 1$ (long dash), $\mu = 2$ (short dash), $\mu = 5$ (chain dash), and $\mu = 10$ (double chain dash), along with the original fluxes (solid): (a) zonal momentum flux, (b) meridional momentum flux, (c) heat flux, and (d) moisture flux.

ing compression rates, so that the overall vertical structure is retained.

The flux reduction may be partially recovered by a ''renormalization'' procedure. The simplest procedure for such ''renormalization'' is to divide the compressed flux profile by the reproducibility measure $\alpha(\mu)$ defined by Eq. (5.3). These ''renormalized'' fluxes are shown in Fig. 14. ''Renormalization'' of compressed fluxes can reproduce the original profiles reasonably well even for a compression with $\mu = 5$ (chain dash). A higher compression with $\mu = 10$ (double chain dash), where less than 0.2% of data is retained, gives less exact results, but still an overall similarity is noted. A further analysis (Figs. 7 and 8 of Yano et al. 2003) shows that this method works well for the whole experiment period, with the ''renormalization'' factor being relatively independent of time.

6. Continuity

Finally, the mass conservation of the compressed system is investigated. Mass continuity is not expected to be satisfied by compression, as the three velocity components are truncated independently. According to Schneider et al. (1997), the mass conservation constraint is one of the major challenges in constructing a wavelet-

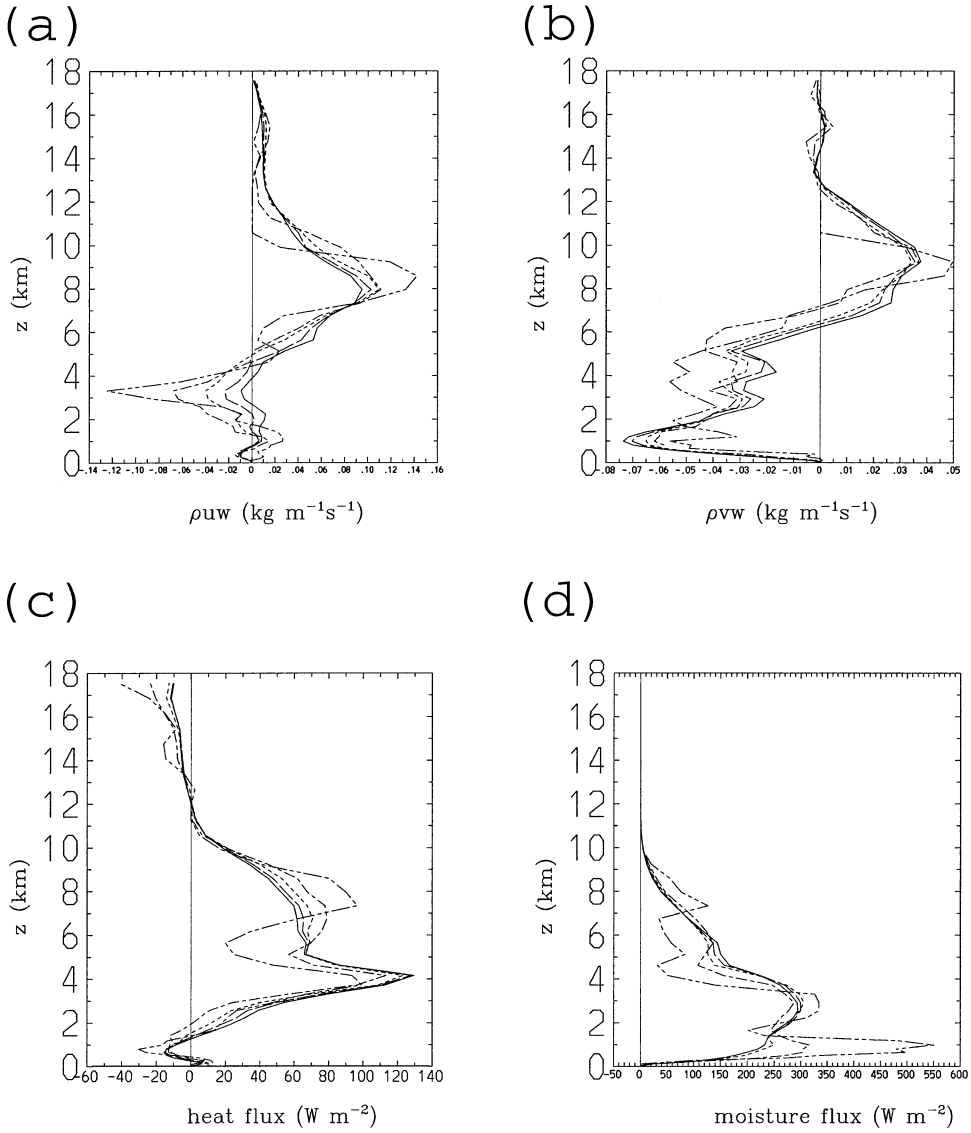

FIG. 14. The same as Fig. 13, but after ''renormalizations'' by a factor defined by Eq. (5.3).

based three-dimensional fluid dynamics model. Methods of divergence-free wavelets have been developed for incompressible unstratified fluids (Urban 1995; Suter 1995). Nevertheless, it would be much easier if standard wavelets could compress the system without substantially deteriorating the mass continuity.

The degree of mass conservation is quantified with the aid of the two quantities

$$\left[\frac{\int (\nabla \cdot \rho \mathbf{v}^{\mu})^2 \, d^3r \int d^3r}{\left(\int \rho \, d^3r\right)^2}\right]^{1/2}, \tag{6.1a}$$

$$\left[\frac{\int \left(\frac{1}{\rho}\nabla \cdot \rho \mathbf{v}^{\mu}\right)^2 d^3r}{\int d^3r}\right]^{1/2}, \tag{6.1b}$$

which are called the mass fluctuation and the divergence error, respectively. The integrals are over the whole three-dimensional-analysis domain. Centered-finite differences are used for evaluating the divergence operators in the above formula, consistent with the original model numerics. In Fig. 15 these quantities are plotted for the Meyer (1), Haar (2), and Daubechies (with $p = 2$) wavelets (3), after normalization by the values ob-

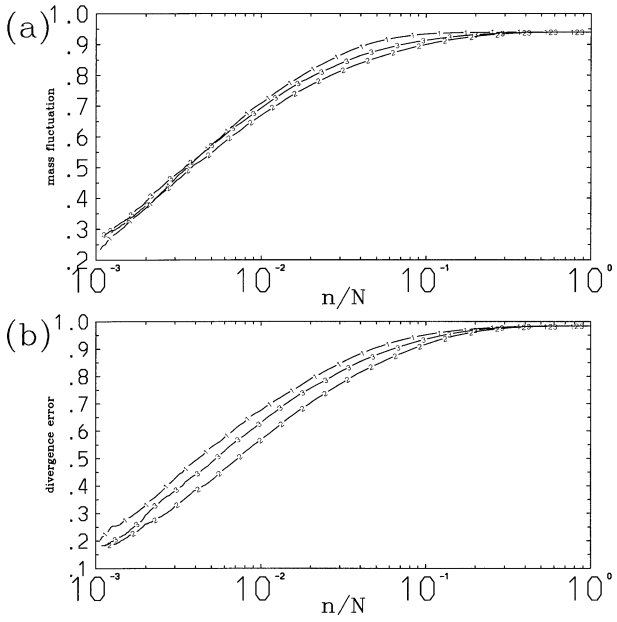

FIG. 15. The accuracy of mass continuity under the normalization by the values of the original data, plotted against the compression rate n/N, is measured by (a) mass fluctuation and (b) divergence error as defined by Eqs. (6.1a) and (6.1b). The curves are for the Meyer (1), Haar (2), and Daubechies wavelets with $p = 2$ (3). The snapshot at $t = 30$ h is used.

tained from the original gridpoint data, which are 1.40×10^{-3} and $4.39 \times 10^{-3}\ s^{-1}$, respectively, for these two quantities. Note that these normalization factors are evaluted offline with the single-precision outputs on staggered grids. These normalization factors (errors) would been smaller if the original double-precision data were used. With both measures, the uncompressed values (with $n/N = 1$) represent the errors resulting after converting the staggered grids into homogeneous grids, as described in section 3.

In spite of the inconsistent compression method, the actual mass conservation of the system is not deteriorated by increasing compression, but simply monotonously improves with both measures. This is presumably because of the decrease of the total variability that contributes to these error measures. The Meyer wavelet produces the highest errors among these three wavelets, but the difference appears to be not significant. This preliminary analysis indicates that the standard wavelets may still be used for the compressed representation of the convective systems without seriously deteriorating the mass continuity.

7. Conclusions

The suitability of wavelets for data compression has been tested using CRM data of convective systems observed during TOGA COARE. In a compressed representation of data, artificial edge wavelike structures generated around pulselike signals are a major concern, in particular for positive-definite physical fields such as moisture and total condensate. This also often leads to unphysical negative local anomalies. The Haar wavelet appears to suffer least from this problem by reproducing the zero-liquid-water region with the least bumps. This aspect makes the Haar wavelet preferable to other wavelets for this application.

However, in spite of vastly different structures for the chosen wavelets, the ability of compression as measured by the explained variance or rms error are overall very similar. Different wavelet modes filter out different structures as ''noises'' depending on their characteristics, and no wavelet clearly outperforms the others concerning the pure data compression problem. This conclusion agrees with the compression experiment on boundary layer data by Katul and Vidakovic (1998).

It is emphasized that the representation problem is essentially a pattern-fitting problem, where a basic set of modes is required that represents the expected physical signal. If the system is interpreted as a set of solitary waves, then the Meyer wavelet serves best for this purpose. On the other hand, if one would like to approximate the system as a sum of idealized pairs of equal-magnitude ''downdraft'' and ''updraft'' with rectangular shapes, the Haar wavelet is most suitable. With the present coding, the Haar wavelet transform also turned out to be roughly a factor of 7 more efficient numerically than the Meyer wavelet transform. In spite of their equally good performance in quantitative measures, the authors feel that the Daubechies wavelets of lower degrees are less appropriate to represent atmospheric convection because of their highly irregular shapes, although they may be useful to represent highly intermittent turbulent structures.

Compressibility depends sensitively on the physical variables. The explained variance decays the slowest for the horizontal wind components and the temperature, maintaining the explained variance above 95% almost down to the 1% compression. On the other hand, for the vertical velocity, the total water mixing ratio, and the total condensate mixing ratio, a compression of no more than 10% can be achieved in order to retain more than 90% of the total variance. However, more than 50% of the total variance for these variables is still retained even with a compression of data to 1%. The compression of the vertical fluxes is found to be rather insensitive to the choice of wavelets for all variables, as over 90% of the total flux is explained by retaining only 10% of the total modes.

Overall, more than 90% of the total variance of the system can be retained with a compression rate of 10%, when the Lorenz threshold with $\mu = 1$ is used. Practically, this leads to a compression of 20% of the original data size, because the wavelet-index parameters must also be stored together with the wavelet coefficients. On the technical level it is assumed that the variable is originally given by and also stored in single precision, and that the index parameters, which are combined into

a single parameter, are also stored as a single-precision integer. The size of the integer parameter might be further compressed by shortening the bit length.

Nevertheless, a loss of variability of the system with a higher compression rate is hard to dismiss. When a compression rate of, say, 1% is pursued, the retained variance typically becomes as low as 50%. Hence, we have to add a certain vertical eddy-flux parameterization in order to compensate the decrease of resolved vertical flux. A simple procedure based on ''renormalization'' has been tested (cf. Yano et al. 2003).

The main concern with compressing three-dimensional meteorological data by wavelets has been their ability to maintain the mass continuity. The present analysis suggests that a straight wavelet compression can maintain the mass continuity of the system reasonably well, although such a constraint is not explicitly imposed.

Finally, the compressed wavelet data can be directly used for various data analyses including budget analyses of heat, moisture, and energy (cf. Fournier 2002, 2003). These types of analyses have been limited mostly to the domain means. The wavelet opens a way to easily extend such analyses for the subdomain scales by focusing on a particular event within the system, for example, the triggering of a convective event by subdomain-scale processes. Although only deep convective data of the tropical atmosphere have been used, the same compression method can readily be applied to large eddy data of boundary layer turbulence or simulations of mesoscale to synoptic scale disturbances by limited-area models. Compression of long observational time series with frequent measurements, as is often the case for wind profilers, is another attractive application. Such a time series often consists of intermittent short time-scale bursts, which can be well retained by the wavelet compression.

Acknowledgments. The authors acknowledge Jean-Philippe Lafore for various discussions, Marie Farge for her encouragements, and Michio Yamada for providing the Meyer wavelet code. Their gratitude goes to two anonymous reviewers whose insightful comments substantially helped to improve the manuscript.

REFERENCES

Attié, J.-L., and P. Durand, 2003: Conditional wavelet technique applied to aircraft data measured in the thermal internal boundary layer during sea-breeze events. *Bound.-Layer Meteor.*, **106,** 359–382.

Chang, S. G., B. Yu, and M. Vetterli, 2002: Adaptive wavelet thresholding for image denoising and compression. *IEEE Trans. Image Process.*, **9,** 1532–1546.

Charney, J. G., 1963: A note on large-scale motions in the Tropics. *J. Atmos. Sci.*, **20,** 607–609.

Daubechies, I., 1994: Two recent results on wavelets: Wavelet bases for the interval, and biorthogonal wavelets diagonalizing the derivative operators. *Recent Advances in Wavelet Analysis,* L. L. Schumaker and G. Webb, Eds., Academic Press, 237–258.

Desrochers, P. R., and S. Y. K. Yee, 1999: Wavelet application for mesocyclone identification in Doppler radar observations. *J. Appl. Meteor.*, **38,** 965–980.

Diedhiou, A., S. Janicot, A. Viltard, P. de Felice, and H. Laurent, 1999: Easterly wave regimes and associated convections over West Africa and the tropical Atlantic: Results from the NCEP/NCAR and ECMWF reanalysis. *Climate Dyn.*, **15,** 795–822.

Donoho, D., and I. Johnstone, 1994: Ideal spatial adaptation by wavelet shrinkage. *Biometrika,* **81,** 425–455.

Ekstrom, P. A., and J. M. Hales, 2000: A wavelet-based approach for atmospheric pollution modeling: Algorithm development. *Mon. Wea. Rev.*, **128,** 3169–3186.

Farge, M., K. Schneider, and N. Kelahan, 1999: Non-Gaussianity and coherent vortex simulation for two-dimensional turbulence using an adaptive orthogonal basis. *Phys. Fluids,* **11,** 2187–2201.

Fournier, A., 2000: Introduction to orthonormal wavelet analysis with shift invariance: Application to observed atmospheric blocking spatial structure. *J. Atmos. Sci.,* **57,** 3856–3880.

——, 2002: Atmospheric energetics in the wavelet domain. Part I: Governing equations and interpretation for idealized flows. *J. Atmos. Sci.,* **59,** 1182–1197.

——, 2003: Atmospheric energetics in the wavelet domain. Part II: Time-averaged observed atmospheric blocking. *J. Atmos. Sci.,* **60,** 319–338.

Fulton, S. R., and W. H. Schubert, 1985: Vertical normal mode transforms: Theory and application. *Mon. Wea. Rev.,* **113,** 647–658.

Galmarini, S., and J.-L. Attié, 2000: Turbulent transport at the thermal internal boundary layer top: Wavelet analysis of aircraft measurements. *Bound.-Layer Meteor.,* **94,** 175–196.

Grabowski, W. W., X. Wu, and M. W. Moncrieff, 1996: Cloud-resolving modeling of tropical cloud systems during Phase III of GATE. Part I: Two-dimensional experiments. *J. Atmos. Sci.,* **53,** 3684–3709.

Grist, J. P., 2002: Easterly waves over Africa. Part I: The seasonal cycle and contrasts between wet and dry years. *Mon. Wea. Rev.,* **130,** 197–211.

Gu, G., and C. Zhang, 2001: A spectrum analysis of synoptic-scale disturbances in the ITCZ. *J. Climate,* **14,** 2725–2739.

Guichard, F., J.-P. Lafore, and J.-L. Redelsperger, 1997: Thermodynamical impact and internal structure of a tropical convective cloud system. *Quart. J. Roy. Meteor. Soc.,* **123,** 2297–2324.

——, J.-L. Redelsperger, and J.-P. Lafore, 2000: Cloud-resolving simulation of convective activity during TOGA-COARE: Sensitivity to external sources of uncertainties. *Quart. J. Roy. Meteor. Soc.,* **126,** 3067–3095.

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

——, ——, and H. Mitsudera, 2002: Scale utilization and optimization from wavelet analysis for data assimilation: SUgOiWADAi. *J. Atmos. Oceanic Technol.,* **19,** 747–758.

Kasahara, A., and K. Puri, 1981: Spectral representation of three-dimensional global data by expansion in normal mode functions. *Mon. Wea. Rev.,* **109,** 37–51.

Katul, G., and B. Vidakovic, 1998: Identification of low-dimensional energy-containing/flux-transporting eddy motion in the atmospheric surface layer using wavelet thresholding methods. *J. Atmos. Sci.,* **55,** 377–389.

Kestin, T. S., D. J. Karoly, J. I. Yano, and N. A. Rayner, 1998: Time–frequency variability of ENSO and stochastic simulations. *J. Climate,* **11,** 2258–2272.

Lafore, J. P., and Coauthors, 1998: The Meso-NH Atmosphere Simulation System. Part I: Adiabatic formulation and control simulations. *Ann. Geophys.,* **16,** 90–109.

Lau, K.-M., and H. Weng, 1995: Climate signal detection using wavelet transform: How to make a time series sing. *Bull. Amer. Meteor. Soc.,* **76,** 2391–2402.

Li, T.-H., 1999: Multiscale representation and analysis of spherical data by spherical wavelets. *SIAM J. Sci. Comput.,* **21,** 924–953.

Luo, J., and L. Jameson, 2002: A wavelet-based technique for iden-

tifying, labeling, and tracking of ocean eddies. *J. Atmos. Oceanic Technol.,* **19,** 381–390.
Mak, M., 1995: Orthogonal wavelet analysis: Interannual variability in the sea surface temperature. *Bull. Amer. Meteor. Soc.,* **76,** 2179–2186.
Mallat, S., 1998: *A Wavelet Tour of Signal Processing*. 2d ed. Academic Press, 637 pp.
Marcal, A. R. S., B. Triebfürst, C. Schneider, and R. A. Vaughan, 2000: Compression of NOAA/AVHRR data with a wavelet transform. *Int. J. Remote Sens.,* **21,** 595–610.
Mesinger, F., and A. Arakawa, 1976: *Numerical Methods Used in Atmospheric Models.* GARP Publication Series, Vol. 14, WMO/ICSU Joint Organizing Committee, 64 pp.
Qiu, J., K. T. PawU, and R. H. Shaw, 1995: The leakage problem of orthonormal wavelet transforms when applied to atmospheric turbulence. *J. Geophys. Res.,* **100,** 25 769–25 779.
Randall, D. A., M. Khairoutdinov, A. Arakawa, and W. Grabowski, 2003: Breaking the cloud parameterization deadlock. *Bull. Amer. Meteor. Soc.,* **84,** 1547–1564.
Sato, K., and M. Yamada, 1994: Vertical structure of atmospheric gravity waves revealed by the wavelet analysis. *J. Geophys. Res.,* **99,** 20 623–20 631.
Schneider, K., N. Kevlahan, and M. Farge, 1997: Comparison of an adaptive wavelet method and nonlinearly filtered pseudospectral methods for two-dimensional turbulence. *Theor. Comput. Fluid Dyn.,* **9,** 191–206.
Schröder, P., and W. Sweldens, 1995: Spherical wavelets: Efficiently representing functions on the sphere. *Comput. Graphics Proc.,* **95,** 161–172.
Suter, D., 1995: Divergent-free wavelets made easy. *Wavelet Applications in Signal and Image Processing III,* A. F. Laine, M. A. Unser, and M. V. Wickerhauser, Eds., SPIE, 102–113.
Torrence, C., and G. P. Compo, 1998: A practical guide to wavelet analysis. *Bull. Amer. Meteor. Soc.,* **79,** 61–78.
Tung, W.-W., and M. Yanai, 2002: Convective momentum transport observed during the TOGA COARE IOP. Part II: Case studies. *J. Atmos. Sci.,* **59,** 2535–2549.
Urban, K., 1995: On divergence-free wavelets. *Adv. Comput. Math.,* **4,** 51–82.
Vasilyev, O. V., D. A. Yuen, and S. Paolucci, 1997: Solving PDEs using wavelets. *Comput. Phys.,* **11,** 429–435.
Weng, H., and K.-M. Lau, 1994: Wavelets, period doubling, and time–frequency localization with application to organization of convection over the tropical western Pacific. *J. Atmos. Sci.,* **51,** 2523–2540.
Xie, S., and Coauthors, 2002: Intercomparison and evaluation of cumulus parameterizations under summertime midlatitude continental conditions. *Quart. J. Roy. Meteor. Soc.,* **128,** 1095–1136.
Yamada, M., and K. Ohkitani, 1991: Orthonormal wavelet analysis of turbulence. *Fluid Dyn. Res.,* **8,** 101–115.
Yano, J.-I., M. W. Moncrieff, and X. Wu, 2001a: Wavelet analysis of simulated tropical convective cloud systems. Part II: Decomposition of convective-scale and mesoscale structure. *J. Atmos. Sci.,* **58,** 868–876.
——, ——, ——, and M. Yamada, 2001b: Wavelet analysis of simulated tropical convective cloud systems. Part I: Basic analysis. *J. Atmos. Sci.,* **58,** 850–867.
——, P. Bechtold, and J.-L. Redelsperger, 2003: ''Renormalization'' approach for subgrid-scale representations. *J. Atmos. Sci.,* **60,** 2029–2038.
——, R. Blender, C. Zhang, and K. Fraedrich, 2004: 1/*f*-Noise and Pulse-like events in the tropical atmospheric surface variabilities. *Quart. J. Roy. Meteor. Soc.,* in press.