Filtering of nonpredictable scales of precipitation can be used to improve forecast precision (rms). Previous papers have studied the scale dependence of predictability of patterns of instantaneous rainfall rate and of probabilistic forecasts. In this paper, motivated by the often localized, intermittent nature of rainfall, the wavelet transform is used to develop measures of predictability at each scale. These measures are then used to design optimal forecast filters. This method is applied to radar composites of rainfall reflectivity over much of the continental United States and is developed to be appropriate for operational forecasts of rainfall rates and raining areas. For the four precipitation events studied, the average correlation at 4-h lead time was increased from 0.50 for the original nowcasts to 0.62 with forecast filtering. This forecast filtering is incorporated into the McGill Algorithm for Precipitation Nowcasting by Lagrangian Extrapolation (MAPLE), which now includes variational echo tracking, a semi-Lagrangian advection scheme, scale-based filtering, and appropriate rescaling of the filtered nowcast fields.

1. Introduction

Weather radar is a crucial tool for the short-term (several hours) quantitative prediction of precipitation. The automated production of such forecasts, called nowcasting, is now widely implemented, and most radar networks are equipped with specialized nowcasting software (Browning et al. 1982; Walton et al. 1985; Seed 2003). In Part I of this series of papers (Germann and Zawadzki 2002, hereinafter Part I), the scale dependence of the predictability of precipitation patterns in continental-scale nowcasts was investigated. In Part II, predictability was studied using probabilistic forecasts developed from nowcasts (Germann and Zawadzki 2004, hereinafter Part II). In this paper we take the additional step of applying measures of scale-dependent predictability to the forecast method itself, designing forecast filters that reduce rms errors. With appropriate transformations of the filtered nowcasts, forecast skill according to the critical success index can also be improved. Because these forecast filters and subsequent transformations are incorporated into the McGill Algorithm for Precipitation Nowcasting by Lagrangian Extrapolation (MAPLE), we begin with a review of the development of nowcasting at McGill.

The weather radar group at McGill has been active in the development of nowcasting algorithms over the past 30 years. Early study of the correlation structure of patterns and echo motion by Zawadzki (1973) was applied in the first nowcasting trials of Austin and Bellon (1974). Useful short-term forecasts were produced by assuming Lagrangian persistence, such that precipitation patterns would continue to move with a single, steady velocity without any growth or decay. The displacement vector was determined by maximizing cross correlation between current and previous observations. The resulting implementation in the short-term precipitation forecasting procedure called SHARP for operational use by forecasters in Canada is described in Bellon and Austin (1978). Since then, the McGill group has worked on several enhancements to the basic method. Bellon et al. (1980) applied nowcasting to geostationary satellite images of clouds in the visible and infrared in a system called RAINSAT. An important negative result was the unsuccessful attempt of Tsonis and Austin (1981) to improve forecasts by accounting for rainfall tendencies. By 1982, the single displacement vector in the original SHARP was replaced by a nine-vector motion field. As a further development, a variational method was developed to derive the field of
motion of radar patterns with any predetermined resolution (Laroche and Zawadzki 1995).

Predictability of Lagrangian persistence forecasts of precipitation was studied by Zawadzki et al. (1994). Predictability of each spatial scale was found to be time dependent and related to synoptic-scale parameters. Selective filtering of perishable scales improved forecast precision (Bellon and Zawadzki 1994). Seed (2003) interpreted these results in the context of a cascading multifractal description of precipitation (Schertzer and Lovejoy 1987) and applied Fourier low-pass filters to improve forecast scores. Part I of the present series of papers also studied the scale dependence of predictability for continental-scale Lagrangian nowcasts using Fourier (discrete cosine transform) low-pass filtering, but without invoking any assumptions of scaling. In Part II, scale dependence for probabilistic Lagrangian nowcasting using a high-resolution field of echo motion was examined. Now, in Part III, we take advantage of the findings of Part I to improve operational nowcasting algorithms. By evaluating predictability at each spatial scale, nowcasts are filtered to reject unpredictable scales and to optimize forecast precision.

Here also, we take an approach that does not require any assumptions of scaling. Section 2 develops a framework for the multiscale analysis of predictability and the design of forecast filters using the wavelet transform. This local transform provides a more appropriate representation of a precipitation field as a sum of localized rainfall pulses at each scale. In section 3 wavelet-based forecast filters are applied to nowcasts based on radar composite observations of precipitation. Optimal and more operationally useful near-optimal forecast filters provide significant gains in forecast precision as measured by correlation or root-mean-square (rms) errors. Appropriate transformation of filtered forecasts also improves skill scores for predictions of rain areas above a certain threshold. Section 4 discusses the potential for these wavelet-based forecast filters, the effects of filtering, and directions for future work.

2. Method

a. Motivation

Precipitation patterns include variability over a wide range of scales. Various groups have sought classification schemes that would be convenient for defining areas of study. One such classification for the mesoscale (Orlanski 1975) is shown in Table 1. A localized, short-lived thunderstorm at the meso-γ-scale may be embedded in a more durable meso-β-scale squall line, which may then be part of a meso-α-scale frontal system that continues for several days. Such classification schemes are somewhat arbitrary and vary between individual authors and research groups.

We cannot predict every detail of the weather, especially precipitation. Which details can we hope to predict? Lorenz (1969) made a theoretical study of the predictability for a fluid, such as the atmosphere, that possesses many scales of motion. He defined the range of predictability as the time until forecast errors exceed some prechosen magnitude. He also noted that errors in predicting a thunderstorm might become large within 15 min, while errors in predicting larger-scale atmospheric features may not become large for days. This is reflected in the time scales mentioned in Table 1.

Of course, the range of predictability (hereinafter referred to simply as the predictability) at each scale will depend on the forecast method being used. Even a climatological prediction has some limited skill, and a persistence forecast may be useful for short-range forecasts. With details on storm motion, a nowcast gives useful predictions for up to several hours. Numerical models containing details of atmospheric dynamics, combined with sufficient measurements as initial conditions, may have some skill up to several days. However, if the same forecast method is applied to various precipitation events, differences in forecast errors will reflect the predictability of each event. In some cases, frontal systems move for many hours over hundreds of kilometers with little change. Nowcasting performs well in these situations. In other cases, a previously stable air mass reaching a strong source of surface heating and moisture may result in rapid storm growth, which will not be predicted by a Lagrangian extrapolation nowcast. With a particular forecast method, some precipitation features are predictable and others are not. Attempts to forecast unpredictable features will only contaminate the skill of the forecast; forecasts that focus on what is predictable will have greater accuracy.

Selective filtering of smaller scales from nowcasts of precipitation can reduce forecast errors. Although the division between which scales are predictable will vary between cases and depend strongly on the forecast lead time, the generally expected pattern is that smaller scales will be less predictable. Bellon and Zawadzki (1994) showed the reduced predictability of smaller scales in precipitation by studying spatially smoothed nowcasts. More recently, in Part I and in work by Seed (2003), Fourier low-pass filtering has been applied for this purpose.

The Fourier transform can be used to represent an image as a combination of waves (sines and cosines) of different wavelengths. Because sines and cosines are extended in space, the Fourier transform is most useful in representing wavelike properties. The forward Four-
The inverse wavelet transform (WT) produces a coefficient $F(k)$ for each wavenumber $k$:

$$f(x) \xrightarrow{\text{WT}} F(k), \quad (1)$$

while the inverse Fourier transform (FT$^{-1}$) reconstructs the original spatial field:

$$F(k) \xrightarrow{\text{FT}^{-1}} f(x). \quad (2)$$

High-wavenumber $k$s correspond to small scales, lower-wavenumber $k$s correspond to larger scales, and $k = 0$ is associated with the average at the largest scale. Fourier low-pass filtering eliminates contributions from wavenumbers larger than some cutoff $k_o$. After performing the forward Fourier transform, coefficients $F(k), k > k_o,$ are set to zero. The modified set of $F(k)$ is then used in the inverse transform. In this way, only low-wavenumber (large scale) features pass through the filter.

However, rainfall fields usually do not have clear, periodic extended patterns. Fields of rainfall rate (and radar reflectivity) regularly show highly localized features with sharp spatial gradients. Fourier low-pass filtering is not appropriate in such circumstances, and results in the Gibbs effect, which adds nonphysical “ripple” features. The Gibbs effect (illustrated later in Fig. A1) is described in greater detail in appendix A.

The wavelet transform provides an alternative scale decomposition in such circumstances. The localized pulse functions used by the wavelet transform are more effective in representing localized, intermittent fields. Instead of the single Fourier transform coefficient $F(k)$ for each Fourier scale $k$, we obtain many wavelet coefficients $W(m, x)$ at each wavelet scale $m$—one for each position $x$. The forward wavelet transform (WT) produces the following coefficients $W(m, x)$:

$$f(x) \xrightarrow{\text{WT}} W(m, x). \quad (3)$$

At a given position in space, $x$, small $m$ will correspond to smaller scales, and large $m$ will correspond to larger scales. In practice, a long-term weighted average $\bar{f}(x)$ describes a (possibly nonzero) mean around each location. The inverse wavelet transform (WT$^{-1}$) reconstructs the original spatial field,

$$W(m, x), \xrightarrow{\text{WT}^{-1}} f(x). \quad (4)$$

A full set of coefficients $W(m, x)$ for all scales and positions is highly redundant. In this paper we use wavelet transform coefficients for a discrete set of wavelet scales $m$. In fact, we use $m$ as a logarithmic scale index related to the wavelet length scale by

$$\text{LengthScale} = 2^n x, \quad (5)$$

where $x$ is the fundamental resolution or pixel length. The finest resolution $x$ corresponds to $m = 0$, length scale $2x$ corresponds to $m = 1$, length scale $4x$, corresponds to $m = 2$, and so on. The scale index $m = 0, 1, 2, 3 \ldots$ refers to the discrete set of wavelet scales used, and is useful in describing and developing the wavelet transform and forecast filters. When showing actual results, the length scale will be used because it conveys physical meaning more directly. A more mathematical introduction to the wavelet transform is found in Burrus et al. (1998) and other similar texts.

A wavelet filter can be designed to selectively weaken or remove less predictable precipitation features from a forecast. Appropriate spatially localized filters are associated with particular wavelet functions. The simple block averaging of Bellon and Zawadzki (1994) is an example of a localized spatial smoothing function that is less susceptible to the Gibbs effect, as shown in appendix A. Block averaging is the two-dimensional smoothing function associated with the Haar wavelet (described in appendix B), and acts as a wavelet low-pass filter.

Of course, the usefulness of forecast filtering depends on the selective weakening and removal of features only at scales that have lost predictability. Using the wavelet transform for scale decomposition of forecast errors, optimal filters can be developed. This is described in the next section, with more details concerning the wavelet transform in appendix B.

### b. Errors and filtering

In this section we will design a forecast filter that is optimized to reduce forecast errors. The loss of predictability for features of a given scale is not abrupt, but rather decreases smoothly in time. Thus, in optimizing the scale filtering we take advantage of whatever predictability each scale has at a given forecast time. A measure of forecast error is established, then expressed as a sum of contributions from different scales. After defining a general forecast-filtering method using wavelet scale weights, the choice of scale weights is directed by the need to reduce error contributions at all scales.

Scale weights will be small for less predictable scales, reducing their influence.

Forecast error is often measured as the mean squared difference between the forecast and later observed values. For a set of observations $I_o$ and a forecast $I_f$ produced at an earlier time, this is

$$\overline{(I_f - I_o)^2} = \overline{I_f^2} - 2\overline{I_f}I_o + I_o^2, \quad (6)$$

where the overbar indicates averaging over the forecast area. Less predictable features of the precipitation will contribute more to the forecast error. We expect smaller scales to be less predictable than larger scales. To verify this, we need to separate the contributions of different scales to the forecast error. Assuming the independence of contributions to the error from different scales, we should then construct a forecast filter that will minimize error contributions from each scale.

In one dimension, the contribution to $\overline{I_f^2}$ from each scale interval $m$ is given by the wavelet spectrum

$$S_f(m) = 2^{-m}[W_f(m, x)]^2, \quad (7)$$
where the subscript f indicates the forecast, as opposed to observations that are denoted by a subscript o. The mean square $\overline{T_f}$ is the sum of these contributions from each scale $m$ and the square of the mean forecast value

$$\overline{T_f} = \overline{T_o}^2 + \sum_m S_f(m). \quad (8)$$

A similar expression can be written for $\overline{T_o}^2$, the mean square of observation values. In the same way, the contribution to $\overline{T_o}^2$ from each scale interval $m$ is given by the wavelet cospectrum:

$$C_{fo}(m) = 2^{-m} W_f(m, x) W_o(m, x), \quad (9)$$

and $\overline{T_o}^2$ can be expressed as

$$\overline{T_o}^2 = \overline{T_f}^2 + \sum_m C_{fo}(m). \quad (10)$$

Both the wavelet spectrum and cospectrum can be generalized for our two-dimensional forecasts and observations. More details on the wavelet spectrum and cospectrum are found in appendix B.

Substituting Eqs. (8) and (10) into Eq. (6), then separating by scale $m$, we have

$$w(m) \approx w(m) \equiv 1,$$

and then taking the inverse wavelet transform:

$$w(m) W_f(m, x), f(x) \rightarrow I_f. \quad (14)$$

If we set $w(m) = 1$ at some scale, the wavelet coefficients for that scale are unchanged. If we set $w(m) = 0$, the wavelet coefficients at that scale will be set to zero. What are the values of $w(m)$ that will minimize the contribution to the forecast error from each scale? Assuming the independence of contributions to the error from different scales, this should minimize the total forecast error.

By Eq. (7), the wavelet spectrum $S_f$ of the filtered forecast $I_f$ is

$$S_f(m) = 2^{-m} [w(m) W_f(m, x)]^2 = [w(m)]^2 S_f(m). \quad (15)$$

where the subscript ff denotes the filtered forecast. Because the filter is applied only to the forecast, the observation spectrum $S_o(m)$ remains unchanged. By Eq. (9) the wavelet cospectrum of the same observations and the filtered forecast $I_f$ is

$$C_{fo}(m) = 2^{-m} W_f(m, x) W_o(m, x) W_f(m, x) \rightarrow [w(m)]^2 C_{fo}(m). \quad (16)$$

Following the same development as for Eq. (12), the contribution of each scale $m$ to the filtered forecast error is

$$Error_f(m) = [w(m)]^2 S_f(m) - 2[w(m)] C_{fo}(m) - S_o(m). \quad (17)$$

Equation (17) can be minimized by setting the first derivative, with respect to $w(m)$, to be zero,

$$\frac{d\text{Error}_f(m)}{d[w(m)]} = 2[w(m)] S_f(m) - 2C_{fo}(m) = 0, \quad (18)$$

and then solving for $w(m)$. The optimal weights will also depend on the lead time $T$ between making a forecast and the verification. For scales with $S_f(m) \neq 0$, the optimal scale weights are given by

$$w(m, T) = \frac{C_{fo}(m)}{S_f(m)}. \quad (19)$$

From here on, the scale weights, optimized to reduce filtered forecast errors at a particular forecast lead time $T$, will be denoted by $w(m, T)$. These optimal scale weights may also vary from one case to another.

For a scale that is completely predictable, forecast patterns will exactly match the observations, so that

$$C_{fo}(m) = S_f(m) \Rightarrow w(m, T) = 1. \quad (20)$$

For a scale that has become completely unpredictable,

$$C_{fo}(m) = 0 \Rightarrow w(m, T) = 0. \quad (21)$$

Scales with intermediate predictability will have optimal weights $w(m)$ between 0 and 1. For all $m$, as forecast lead time $T \rightarrow 0$,

$$w(m, T \rightarrow 0) = 1, \quad (22)$$

assuming that the observations are compared with themselves because no forecast is required for the time of observation. As $T \rightarrow \infty$, the forecast loses skill and $w(m, T)$ decreases. A long forecast lead time renders the forecast (on average) no better than climatology. For intermediate times, $w(m, T)$ will usually increase with increasing $m$, decrease with increasing $T$, and take on values between 0 and 1.

Appendix B provides more details on the implementation of these scale weights using a wavelet’s associated smoothing function.
3. Observations and results

a. Composite radar data

The U.S. national radar composites used in this analysis are a Weather Services International (WSI) Corporation NOWrad product that has benefited from three levels of quality control. The original resolution is 15 min in time, about 2 km in space (for east–west distances, there is some dependence on latitude), and 5-dBZ intervals of reflectivity. For the analysis presented in this paper the spatial resolution has been reduced to about 4 km, with latitude dependence of east–west distances ignored. Averaging was done in units of dBZ. The spatial coverage extends from 20° to 52°N latitude, and from 130° to 60°W longitude. Reflectivity corresponds to the maximum value measured in the vertical over each pixel by any Weather Surveillance Radar-1988 Doppler (WSR-88D) during the time period.

The accuracy of these reflectivity observations is not comparable to that of rainfall maps based on sophisticated algorithms that include corrections for the vertical profile of reflectivity, visibility, attenuation, $Z-R$ variations, and others. The images certainly suffer from brightband contamination and lack of visibility, which reduces the quality in terms of quantitative rainfall estimates. However, the spatial coverage of the composites is unique, making the product particularly suitable for scale-dependent analyses of predictability up to near-continental scales.

As in Part I, four precipitation events in a subdomain of 2720 km × 2720 km are used to illustrate the development of optimal forecast filters using wavelet statistics. Figure 1 shows one representative image of each system, and Table 2 lists the duration, the spatial extension, and two summary statistics of the rainfall intensity distribution. The criteria for selecting the events was a minimum extension of about $3 \times 10^5$ km$^2$ and sufficient data quality. Some further description of these four cases may be found in Part I.

b. Motion estimation and nowcasts

Motion estimation by variational echo tracking (VET) was presented by Laroche and Zawadzki (1994, 1995). The technique was originally developed to retrieve the three-dimensional wind field from single-Doppler clear-air echoes. It was adapted to continental-scale radar composite images of precipitation with no Doppler information in Part I.

The echo motion field in this study is retrieved in two runs of the motion estimation algorithm. A first run produces an initial coarse guess on a $5 \times 5$ grid (of 520 km × 520 km squares). This is then used to initialize the subsequent minimization on a finer $25 \times 25$ grid (of 104 km × 104 km squares).

The extreme variability of reflectivity fields at the high resolutions can make VET computations slow and unreliable. The VET algorithm may try to match unrelated, intense short-lived thunderstorms if there is insufficient spatial smoothing. If the time step between images being compared is too short, the precipitation may not have moved enough distance to be properly estimated. These problems are mitigated by appropriate spatial smoothing and longer time steps between input images. For motion estimation, the coarse $5 \times 5$ grid estimate was produced using three reflectivity fields 2 h apart (4, 2, and 0 h before the chosen time), smoothed to 64 km × 64 km. With the coarse-grid motion estimate as a reasonable initial guess to stabilize the VET algorithm, the final computations on a $25 \times 25$ grid used three reflectivity fields 30 min apart (60, 30, and 0 min before the chosen time), smoothed to 16 km × 16 km. Experimentation showed that these parameters permit-

### Table 2. Statistics of the four precipitation events in a subdomain

<table>
<thead>
<tr>
<th>Date</th>
<th>Start (UTC)</th>
<th>Duration (h)</th>
<th>Extension ($10^5$ km$^2$)</th>
<th>$&gt;1$ mm h$^{-1}$ (%)</th>
<th>$&gt;10$ mm h$^{-1}$ (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>30 Jul 1998</td>
<td>0400</td>
<td>16</td>
<td>6.0</td>
<td>42</td>
<td>6.4</td>
</tr>
<tr>
<td>1 Jun 1999</td>
<td>1600</td>
<td>25</td>
<td>3.6</td>
<td>47</td>
<td>12</td>
</tr>
<tr>
<td>16 May 2000</td>
<td>0000</td>
<td>22</td>
<td>3.9</td>
<td>34</td>
<td>2.0</td>
</tr>
<tr>
<td>25 May 2001</td>
<td>0400</td>
<td>20</td>
<td>6.1</td>
<td>33</td>
<td>2.6</td>
</tr>
</tbody>
</table>
ted the estimation of echo motion with sufficient spatial detail, consistency between subsequent time periods, and sufficient computational stability. The time periods for the finer-grid calculations (including observations for 1 h before chosen times) are listed in Table 2.

Nowcasts were produced with the semi-Lagrangian backward scheme of Part I. As in that study, a forecast lifetime can be defined using the correlation $c(T)$:

$$c(T) = \frac{\tilde{T}_f}{\sqrt{T_o^2 T_f}},$$

(23)

such that

$$\text{Lifetime} = \int_0^\infty c(T) \, dT.$$  

(24)

For an exponential correlation function, the lifetime is the time for which the correlation drops to $(1/e)$. Recall that the mean value of our observations $\bar{T}_o$ and the mean value of the forecast $\bar{T}_f$ are usually nonzero. For example, the average rainfall rate over an area with any rain will be nonzero and positive. In our definition of the correlation and lifetime, we have not subtracted mean values from the observations and forecast, because zero rainfall is the only meaningful reference value. When using radar reflectivities, a reference value for zero rainfall must be chosen that is below the minimum detectable rain threshold. In this paper, the value of 0 dBZ is assigned to nonraining areas.

This definition of lifetime was applied to forecast sets with up to 8 h lead time for the four precipitation cases. The correlation function was assumed to be approximately exponential. The increased predictability of larger-scale features is apparent in Fig. 2, where lifetimes increase dramatically with increasing cutoff scale for either Fourier or wavelet low-pass filters. In each case, the same low-pass filter is applied to the forecasts and the observations used for comparison. Removing smaller, more perishable scales increases lifetime values. This differs in certain ways from a similar analysis in Part I. There, Fourier low-pass-filtered observations were used as input to the nowcasting method. Here, the filtering is applied to the output of the nowcasting routine. More important, Part I sought to avoid effects of the ripples due to the Gibbs effect (as shown in Fig. A1b) by applying a threshold of 10 dBZ to both the low-pass-filtered forecasts and the observations used for comparison. With increasing smoothing, greater areas of reflectivity are reduced below this threshold, dramatically reducing correlation and lifetime values.

As explained in appendix A, Haar wavelet low-pass filter results can be compared with Fourier low-pass filter results if a $\sqrt{2}$ correction factor is used to determine an equivalent Fourier cutoff scale. This equivalent cutoff scale is used in showing wavelet low-pass lifetimes in Fig. 2. The results for Fourier and wavelet low-pass filtering are very similar. However, with wavelet low-pass filtering there is no need to apply arbitrary thresholding. With Fourier low-pass filters, such thresholding is necessary to eliminate widespread ripples due to the Gibbs effect, but this reduces forecast lifetimes in a manner that depends on the threshold used.

As found in Part I, the cases of 16 May 2000 and 30 July 1998 are most predictable. Forecast lifetimes are shortest for 1 June 1999, and the case of 25 May 2001 shows an intermediate predictability. The curves in Fig. 2 are roughly linear, with similar slopes between the four cases. Very few parameters may ultimately be required to describe the scale dependence of predictability.

c. Spectra and cospectra

The scale weights $w(m, T)$ for an optimal forecast filter (OFF) will depend on the wavelet spectrum and cospectrum, as described in Eq. (19). Here we present averaged Haar wavelet spectra and cospectra from the four precipitation events.

Figure 3 shows the average Haar wavelet spectrum of logarithmic reflectivity (dBZ) for observations of the 25 May 2001 case. Spectral values are normalized by the mean square dBZ level, with larger scales to the left and smaller scales to the right. The average (solid line) curve in Fig. 3 was produced by averaging spectra from
all of the 1-h intervals during the period. The values for these individual spectra are also plotted to show the magnitude of variations with time. The spectral peak at scales of about 250 km indicates the approximate size of the dominant precipitation patterns. This corresponds to the width of the strong rainbands that are part of a larger comma-shaped storm organization that contributes to spectral values at larger scales. Within individual rainbands, a variety of more local storms contribute to spectral values at smaller scales.

Figure 4 compares the average spectra for each of the four cases studied. In each case, the spectrum decreases with decreasing scale for scales less than about 150 km. The slopes of the spectra are similar in this range, suggesting that some aspects of precipitation organization at these scales are independent of the larger-scale forcing. The greatest differences between the four cases are seen for scales greater than 250 km. As previously noted with the case of 25 May 2001, precipitation is compressed in localized rainbands, giving a spectral peak near 250 km. For the case of 30 July 1998, the spectral peak is at scales closer to 500 km. In contrast, the spectrum for the 16 May 2000 case with more widespread and uniform precipitation continues to increase with scale length.

Of course, the shape of these spectra would be different if calculated using rainfall rates or linear reflectivity instead of dBZ values. The logarithmic transformation from rainfall rate or linear reflectivity to logarithmic reflectivity (dBZ) reduces the relative influence of strong, localized peaks in reflectivity on the spectrum. The strongest rainfall rates are often found in isolated, small-scale convective cells. This will affect both the spectra and cospectra, reducing the values at smaller scales. This, in turn, will affect the optimal scale weights $w(m, T)$ for our forecast filter. The optimal forecast filter is designed to minimize the mean square error in whatever units are used. Here, logarithmic reflectivity units (dBZ) are used.

Returning to the specific case of 25 May 2001, Fig. 5 shows wavelet cospectra between predicted and observed reflectivity measurements for different forecast lead times. Cospectra in this figure are normalized by the product of rms of the forecast and observations. Treating the observations as the “prediction” at zero lead time, we show the spectrum for the initial $T = 0$ min. Cospectral values at each scale decrease with increasing forecast lead time. The cospectral values decrease most rapidly at smaller scales. For instance, less than 1-h storms, with length scales less than 20 km, have advanced through their life cycle. The Lagrangian nowcast of such storms for 1 h or more has negligible skill. Those features in the nowcast may contribute significantly to the spectrum, but will not contribute much to the cospectrum. As described in Table 1, larger storm patterns evolve more slowly, and are predictable for longer times. Cospectral values for scales greater than...
200 km continue at over one-half of their $T = 0$ min value for as long as 4 h, indicating that larger precipitation patterns were predicted with reasonable accuracy over such times.

d. Optimal and near-optimal forecast filters

Using spectra and cospectra, such as those shown in Figs. 4 and 5, OFF can be derived a posteriori using Eq. (19). The scale weights $w(m, T)$ for the 30 July 1998 and 25 May 2001 cases are shown as solid curves in Fig. 6. After verifying forecasts against later observations, these would have been the optimal scale weights for forecast filtering for different forecast lead times. There are differences in optimal scale weights between cases.

Some a priori approximation of optimal scale weights $w(m, T)$ is required for an operational application to forecasts. It is not enough to know the best forecast filter after the forecast has been verified! An appropriate filter, as close to the best as possible, must be determined and applied at the time that the forecast is made. With sufficient analysis, it might be possible to determine likely filters for a geographical region during different seasons. Here we will derive a “near optimal” forecast filter (NOFF) from scale weights for a single lead time, using a forecast shortly before the time of interest.

We start by taking a 1-h forecast produced 1 h before the chosen “current” time. From this forecast and current observations, a set of recent scale weights $w(m, T)$ can be determined for the OFF at 1-h lead time. These weights, calculated for the previous hour, are taken as near-optimal estimates for the current 1-h forecast. As an example, suppose the current time is 0800 UTC. We first generate a 1-h lead time nowcast using the observations and motion field that were available at 0700 UTC. Comparing the forecast produced at 0700 for 0800 UTC with the current observations (at 0800 UTC) allows us to compute a posteriori smoothing scale weights for a 1-h lead time forecast. We take these weights as still valid for the current forecast (produced at 0800 UTC) for 0900 UTC.

Initial analysis of the decrease in scale weights $w(m, T)$, such as those shown in Fig. 6, suggested a linear transformation of weights between scale and time. This was taken as a means of obtaining reasonable scale weights for a NOFF based on available observations and a recent forecast. For instance, the predictability of features with a spatial scale of 100 km at 2 h may be similar to the predictability of 50-km features at the 1-h lead time. That is, we assume that

$$w(m, 2h) = w(m - 1, 1h).$$

More generally, we produce a full set of NOFF scale weights for different forecast lead times, as follows:

$$w(m, T) = w(m - \log_2 T, 1h),$$

where the forecast lead time $T$ is measured in hours. Because $m$ is related to the length scale, as described in Eq. (5), exponentiating the argument $(m - \log_2 T)$ from the right-hand side of Eq. (26) gives

$$e^{m-\log_2 T} = \text{LengthScale} \frac{T}{T},$$

with units of distance divided by time. This is the rate at which scale weights shift to larger length scales with increasing forecast lead time. It is the parameter used in estimating a linear transformation of weights between scale and time.

As an initial test of Eq. (26), it was applied to the OFF scale weights (solid curves) for the 1-h forecast lead time in Figs. 6a and 6b for the cases of 30 July 1998 and 25 May 2001, respectively. The resulting NOFF scale weights for forecast lead times of 30 min, 2 h, and 4 h are shown as dashed lines. Equation (26) produced excellent estimates of OFF scale weights. The results for 25 May 2001 are more typical of the other cases, with overestimation of scale weights at longer forecast lead times and larger scales. Clearly, Eq. (26) does not always provide accurate scale weights. In Fig. 6b, OFF scale weights decrease more rapidly with fore-
cast lead time than expected by the NOFF scale-weight estimates. The exact form of the scale transformation of weights varies from one case to another. However, the NOFF approximation produces an a priori estimate of the best scale weights with a minimum of case-dependent information.

For forecast lead times that are not integral powers of 1 h, scale weights must be interpolated between the closest available lead times. A linear interpolation of scale weights was used in this study. A logarithmic interpolation of scale weights from nearby lead times would be an alternative.

We now have two smoothing methods designed to reduce forecast errors. The a posteriori OFF gives the best possible reduction of errors. NOFF produces only a rough estimate of what will be the best scale weights. However, because it is based on earlier forecast results and requires a minimum of case-dependent information, it can be applied to forecasts as they are produced.

e. Forecasts and error measures

OFF was designed to minimize the contribution of each filtered forecast scale to the mean squared error. NOFF allows an estimate of the optimal scale weights operationally as the forecasts are made. The result should be filtered forecasts that significantly improve forecast accuracy, according to measures such as the rms error and the correlation of the forecast with observations. In this section we test how well these objectives are met for the four precipitation events listed in Table 2.

The effect of filtering on the appearance of forecast maps in dBZ units is shown in Fig. 7. Figure 7a shows observations at 0500 UTC on 25 May 2001. NOFF produced from these observations are shown for the 30-min forecast (Fig. 7b), the 2-h forecast (Fig. 7c) and the 8-h forecast (Fig. 7d). As previously described, details for each scale are progressively removed as they become less predictable. In Fig. 7b, the effects of filtering are not noticeable. With a 30-min forecast, scales greater than tens of kilometers are still largely predictable. Average scale weights for the 25 May 2001 case are shown in Fig. 6b. NOFF smoothing of scales of tens of kilometers is not really visible on a map of this scale that covers a distance of thousands of kilometers. The effects of forecast filtering begin to be visible in the 2-h NOFF results, shown in Fig. 7c. For this forecast lead time, even scales near 100 km suffer a significant loss of predictability and are affected by NOFF smoothing. These scales are large enough to be seen on this map. For the 8-h forecast, shown in Fig. 7d, the dramatic effects of filtering at scales up to hundreds of kilometers are seen. There are two related effects. First, the forecast area of (nonzero) precipitation is greater, because the exact future location of the rainfall is uncertain. Second, the average forecast rainfall rate over raining areas is reduced as the precipitation is smoothed over larger regions.

As the lead time \( T \to \infty \), these forecast filters will average the predicted precipitation to a mean value for the entire area. This gives something like a climatological estimate, which is not really useful as a nowcast. For each precipitation case there will be a forecast lead time beyond which the filtered nowcast is of no practical value. This visible loss of forecast clarity can be initially disturbing. How can we expect to improve forecast error statistics by removing some of the detail? Figures 8a–8d show the rms errors of original and (both OFF- and NOFF-) filtered forecasts as a function of forecast lead time. The rms values, in units of dBZ, were calculated using only locations with either a forecast or an observed intensity greater than 2 dBZ. The filtered forecasts have lower rms errors than the unsmoothed forecasts, even for a 15-min lead time. Filtering has selectively removed the generally inaccurate, small-scale details. The improvement becomes greater for longer lead times. However, rms error increases with lead time for both the original and filtered forecasts. OFF usually has lower rms errors than NOFF. The two only begin to differ by more than 1 dBZ for forecast lead times greater than a few hours. This is important because Fig. 6b showed some differences between average NOFF and OFF scale weights for the 25 May 2001 case. Apparently the filtered forecast’s rms error is not very sensitive to these minor differences in scale weights. The near-optimal forecast filtering, which can be applied operationally at the time that forecasts are produced, is a suitable filtering method. At lead times greater than about 6 h, the rms errors for OFF and NOFF may actually decrease, but by this point the smoothing effects of forecast filtering have had dramatic effects on the forecast’s appearance. Forecast correlation with observations is also strongly improved by filtering. The correlation results for the four precipitation events are shown in Figs. 9a–9d. Higher correlation of the original forecasts with observations indicates the relative predictability of the different cases. The event of 1 June 1999 was least predictable, while the event of 16 May 2000 was most predictable. Forecast filtering by both OFF and NOFF improves correlation results in all four cases. Again, the NOFF method provides almost as much improvement as OFF.

These cases suggest that an operational application of forecast filtering with NOFF will reliably improve correlation and reduce rms errors. Differences from OFF results are not significant, and so it is reasonable to focus instead on the effects of implementing NOFF.

f. Smoothing and skill scores

Because OFF and NOFF were designed to improve forecast correlation and rms errors, their skill in predicting areas of rainfall exceeding a threshold must be verified. In Fig. 7, we have already seen that NOFF
forecasts contain less information on intensity distributions as the forecast lead time increases. The total area for which precipitation is predicted increases with increasing forecast lead time. Predicted rainfall spreads over larger areas, compensating for the increasing uncertainty in the future position of precipitation patterns. In this respect, some smoothing is good. Coupled with this, the distribution of reflectivity or rainfall rates becomes more uniform, tending toward lower values. This aspect of smoothing is not good, because it misrepresents the type of rainfall that is expected. Understanding these effects of smoothing allows us to transform the NOFF forecast values for better prediction of forecast areas, with some smoothing but not too much.

For filtered forecasts, there is an apparent contradiction between improvements in error measures and the shift toward less intense, more widely spread areas of forecast precipitation. In fact, the unsmoothed and filtered forecasts are different types of estimates. The unsmoothed forecast represents one of many possible sets of future observations. Forecasts filtered by OFF or NOFF give an average of the various most likely future observations. Such a map of rainfall, averaging over possible outcomes, does not usually resemble any likely
Fig. 8. Effects of OFF and NOFF forecast filters on rms errors for the four precipitation events. The rms errors (dBZ) were calculated for locations where intensities greater than 2 dBZ were forecast and/or observed.

Fig. 9. Effects of OFF and NOFF forecast filters on $c(T)$ correlation for the four precipitation events. For the four cases, the average forecast correlation with 4-h lead time is 0.50 with no smoothing but increases to 0.62 for the NOFF forecasts.

individual future rainfall field. Although OFF and NOFF were designed to reduce rms errors, this does not guarantee that all other forecast error measures will be improved. The spread of the precipitation area due to filtering may be beneficial to the extent that it captures the uncertainty in the future position of precipitation. Alternatively, this spread could be optimized with respect to “hits” and “misses” of precipitation.

The critical success index (CSI) skill statistic is used here to study the ability of a filtered forecast to predict areas with intensities above a threshold. Any image grid point that is correctly forecast to be above the threshold counts as a hit. If a location is incorrectly forecast to be below the threshold, this is a miss. A location that is incorrectly forecast to be above the threshold is a “false alarm.” After determining the total number of hits ($H$), misses ($M$), and false alarms ($F$) over all forecast grid points, the CSI is calculated as

$$\text{CSI}(\mathcal{R}_t) = \frac{H}{H + M + F},$$  \hspace{1cm} (28)

where $\mathcal{R}_t$ is the chosen threshold of reflectivity or rainfall rate. Results of the CSI will depend strongly on the threshold chosen.

The uncertainty in predicting areas with rainfall rate above a threshold increases with forecast lead time, but the net effect of forecast filtering on this uncertainty varies. For example, by spreading the forecast precipitation over a wider area, filtering may increase the number of hits and decrease the number of misses for lower thresholds, although it may also increase the number of false alarms. As well, the total number of locations forecast above a higher threshold decreases sharply with filtering as isolated strong features are smoothed to lower values over a wider area. In this situation, the number of hits and false alarms decreases, and the number of misses increases.

In Fig. 10, effects of OFF and NOFF on the CSI are shown for a 10-dBZ threshold. Naturally, the CSI decreases with increasing lead time whether the original forecast (no smoothing), OFF, or NOFF are used. In addition, for the first few hours of forecast lead time, either OFF or NOFF actually produce a small improvement in CSI for each of the four cases. However, for each case there is a forecast lead time beyond which OFF and NOFF forecasts provide a poorer prediction.
of the area exceeding 10 dBZ than the original forecast with no smoothing. For example, this crossover time is around 8 h for the case of 30 July 1998, while for the case of 25 May 2001 it is about 3 h. In Fig. 11, effects of OFF and NOFF on the CSI are shown for a 15-dBZ threshold. Again, both OFF and NOFF provide a slight increase in CSI for the smaller lead times. For longer lead times the CSI is worse than the original forecast with no smoothing. The crossover time for each case occurs earlier with the 15-dBZ threshold than with the 10-dBZ threshold.

How can OFF and NOFF dramatically reduce forecast rms errors and increase correlation, as shown in Figs. 8 and 9, respectively, while having so little effect on the CSI at the 10- and 15-dBZ thresholds? This apparent contradiction forces us to reexamine the differences between error measures and what is meant by forecast quality. The CSI is based on a small number of categories for a forecast at each location. All hits are considered equally important, and a near miss has the same significance as any other miss. By contrast, a near miss will contribute less to the rms error than an average miss. The rms error measure, therefore, rewards the avoidance of large errors. NOFF and OFF are, therefore, designed to avoid large errors, so that as lead time increases and the exact location of more intense storms is more uncertain, increased smoothing avoids large errors by predicting lower intensities over a wider area. The same filtering that reduces rms errors also decreases the CSI for higher reflectivities. As seen in Figs. 10 and 11, this eventually affects even the CSI at the 10- and 15-dBZ thresholds.

Even if the future location of local, intense storms is uncertain, in most cases a forecaster will still need to know that such storms are likely to occur. These details are gradually lost with increasing lead time in NOFF forecasts, such as those shown in Fig. 7. Any modification of the filtered forecast is likely to increase the rms error, but this may be justifiable if the forecast precipitation areas and distributions of forecast values become more realistic. The NOFF forecast “spread” with increasing smoothing can be studied by matching the area exceeding a certain reflectivity threshold. In Fig. 12, this type of matching is done for a set of forecasts from each of the four precipitation events. In Fig. 12a, the area exceeding 15 dBZ is first calculated for each observation image. For the corresponding NOFF (smoothed) forecast, a second reflectivity threshold is found with the same total area exceeding that value. As a result of filtering and spreading, the corresponding
Fig. 12. NOFF forecast thresholds with an area of greater reflectivity equal to the corresponding area exceeding the 15-dB Z threshold in (a) observations and (b) the original, unsmoothed nowcast intensity. Results are shown for forecasts of 0500 UTC 30 Jul 1998, 1700 UTC 1 Jun 1999, 0100 UTC 16 May 2000, and 0500 UTC 25 May 2001.

Table 3. Effects of several methods for rescaled NOFF forecasts on rms error and the CSI at the 15-dB Z threshold.

<table>
<thead>
<tr>
<th>Forecast measure</th>
<th>Original NOFF</th>
<th>S-PROG CSI-max</th>
<th>Area matching</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>2-h rms (dBZ)</td>
<td>4-h rms (dBZ)</td>
<td>8-h rms (dBZ)</td>
</tr>
<tr>
<td>2-h CSI at 15 dBZ</td>
<td>0.325</td>
<td>0.371</td>
<td>0.356</td>
</tr>
<tr>
<td>4-h CSI at 15 dBZ</td>
<td>0.229</td>
<td>0.258</td>
<td>0.232</td>
</tr>
<tr>
<td>8-h CSI at 15 dBZ</td>
<td>0.109</td>
<td>0.125</td>
<td>0.060</td>
</tr>
</tbody>
</table>

NOFF thresholds were less than 15 dBZ. In each case, as forecast lead time and NOFF smoothing increased, these matching NOFF thresholds decreased. The exact decrease with forecast lead time was different for each case. In Fig. 12b, the area exceeding 15 dBZ in each of the original, unsmoothed nowcasts is first calculated. The size of this area matched the area in the NOFF forecast exceeding some other, lower threshold. Decreases in matching NOFF threshold values were again found with increasing forecast lead time. However, the decreases in matching thresholds in Fig. 12b were more organized than those found for the same cases in Fig. 12a. The matching thresholds for the four cases were more similar, and showed roughly linear decreases with the logarithm of forecast lead time.

Results such as those in Fig. 12a can only be determined after forecasts have been verified against observations. The results in Fig. 12b, however, can be determined when forecasts are produced. These area-matching thresholds, reflecting the spatial characteristics of the forecast rainfall patterns, provide one basis for methods of “rescaling” NOFF forecast intensities that are suitable for operational implementation. Here we suggest three possible forecast-rescaling methods, and apply them on sample 2-, 4-, and 8-h lead time nowcasts from 0500 UTC on 26 May 2001. The relative effects of each method on rms error and the CSI for a 15-dB Z threshold are shown in Table 3. These statistics are first given for the original nowcast and for the NOFF version.

In the Spectral Prognosis (S-PROG) model of Seed (2003), a rescaling method is applied to Fourier-filtered nowcasts. The S-PROG rescaling is applied here at each grid point of a NOFF forecast:

$$\text{dBZ(S-PROG rescaled)} = \begin{cases} \text{dBZ(NOFF)} + (15 - Z) & \text{if dBZ(NOFF)} > Z \\ 0 & \text{otherwise}, \end{cases}$$

where $Z$ is the NOFF forecast threshold with the same area exceeding it as the area exceeding 15 dBZ in the original, unfiltered nowcast. This rescaling counters some smoothing effects of filtering, removing widely spread areas of low intensity, while uniformly increasing higher intensities by the same amount to be 15 dBZ and greater. By reversing the spread of forecast areas of precipitation, the CSI for a threshold of 15 dBZ is higher than for NOFF. For the forecasts used in Table 3, the CSI values using S-PROG rescaling are actually greater than the values for the original nowcasts. This improvement in CSI comes at a cost of increasing the rms errors, although not to the level of the original nowcasts. As in section 3e, rms values (dBZ) are calculated using only locations with either a forecast or observed intensity greater than 2 dBZ.

Another rescaling method is based on first finding the single threshold in the original nowcasts and NOFF forecasts that has the maximum CSI. For the example, nowcasts used to produce Table 3, the CSI-maximizing (CSI max) thresholds were very low: 3.70 dBZ for 2-h lead time, 3.50 dBZ for 4 h, and 2.30 dBZ for 8 h. As for the S-PROG rescaling, forecast grid points with NOFF values less than the threshold are set to zero. Intensities above the CSI-max threshold are then increased by a constant value to preserve the average intensity over the entire forecast area. Table 3 shows the effects of CSI-max rescaling to be similar to S-
expression of this uncertainty is the probability forecast plays a single value at each location. A more complete here are based on this principle, each nowcast still dis-
small areas). Although the filtered forecasts presented
less reliable in predicting rainfall rate at a point (or over
rms errors. Where rms error is not the only relevant
accurate, NOFF removes these details in order to reduce
information about smaller storm structures becomes in-
are averaged with adjacent areas. In this way, as forecast
patterns is lost as small regions of higher rainfall rates
low values over large areas. The texture of precipitation
intensities. Thus, they are understandably smoothed to
accurate, NOFF removes these details in order to reduce
rms errors. Optimal forecast filters (OFF) and opera-
tionally useful near-optimal forecast filters (NOFF)
have been developed using measures of predictability
as a function of scale based on the wavelet transform.
These methods were applied to continental-scale La-
grangian nowcasts based on composite radar observa-
tions of precipitation. The forecast filters, which adapt
to changing conditions, provide significant improve-
ments in forecast precision in terms of correlation and
rms errors. For the four precipitation events studied, the
average correlation at the 4-h lead time was increased
from 0.50 for the original nowcasts to 0.62 with forecast
filtering.

Filtered forecasts for longer lead times reflect the lack
of predictability of small scales associated with higher
intensities. Thus, they are understandably smoothed to
low values over large areas. The texture of precipitation
patterns is lost as small areas of higher rainfall rates
are averaged with adjacent areas. In this way, as forecast
information about smaller storm structures becomes in-
accurate, NOFF removes these details in order to reduce
rms errors. Where rms error is not the only relevant
forecast error measure, NOFF forecasts can be rescaled
to better predict rain areas above a threshold at the cost
of some increase of rms errors.

With increasing forecast lead time, forecasts become
less reliable in predicting rainfall rate at a point (or over
smaller areas). Although the filtered forecasts presented
here are based on this principle, each nowcast still dis-
plays a single value at each location. A more complete
expression of this uncertainty is the probability forecast
for rainfall rates at each location described in Part II. Such
products convey more information than filtered forecasts for longer lead times.

The NOFF method should now be applied to a long,
continuous set of nowcasts and verified against observ-
ations. The same tests could also include further cal-
bilation of NOFF rescaling factors for better rain-area
estimates. For hydrological applications, requiring pre-
cipitation estimates over larger areas, forecast filtering
will be helpful. In other situations, it can be used along
with the original nowcast to illustrate forecast uncer-
tainty.

Acknowledgments. This work was funded by a grant
from the Canadian Foundation for Climate and Atmo-
spheric Sciences.

APPENDIX A

The Gibbs Effect

Rainfall fields routinely exhibit highly localized fea-
tures and sharp spatial gradients. The Fourier represen-
tation, with its extended basis functions, has problems
with such behavior. The Gibbs phenomenon occurs for
a Fourier series approximation of a piecewise smooth
function with discontinuities. At each discontinuity, the
Fourier approximation will either overshoot or under-
shoot the actual magnitude of the discontinuity by about
9% (Walker 1988, 60–63). This is accompanied by the
Gibbs effect—a “ringing” or “ripple” extending away
from the discontinuity. The frequency of these ripples
is equal to the first term omitted from the Fourier ap-
proximation (Weaver 1989, 125–127). Fourier low-pass
filters can be modified by various windowing functions
to reduce (but not eliminate) the Gibbs effect (Ersoy

Although the Fourier and wavelet representations dif-
fer, one can compare Fourier and wavelet low-pass fil-
ters. The wavelet spectrum is equivalent to a smoothed
Fourier spectrum (Hudgins et al. 1993), where the
smoothing depends on the wavelet used. Similarly, the
Fourier low-pass cutoff frequency for best comparison
with a wavelet low-pass filter depends on the wavelet
basis function used. Although not well localized in fre-
quency, the power spectrum of the fundamental Haar
wavelet (of length 1) has its main peak centered near a
frequency of 0.74. This corresponds to a Fourier mode
with a wavelength of about 1.35 times the length of the
wavelet. For instance, a Haar wavelet of length 128 km
has a Fourier spectrum with strong components between
spatial scales of 128 and 256 km. Here, we will apply
(to the nearest pixel length) a \( \sqrt{2} \) factor between Haar
wavelet low-pass and Fourier low-pass cutoff frequen-
cies to render the results more comparable. A wavelet
low-pass filter with its scale of 128 km will be compared
with a Fourier low-pass filter with a cutoff wavelength
Figure A1a shows a radar reflectivity field (in units of dBZ) for 0800 UTC on 25 May 2001. In Fig. A1b, a Fourier series approximation is shown, which omits Fourier components for spatial scales less than 180 km × 180 km. The Gibbs effect appears as ripples that extend well away from the discontinuities. Note that white areas of these ripples take on values below the original reflectivity threshold. If the same type of Fourier low-pass filter is applied to the image, expressed as rainfall rates, nonphysical areas of “negative rainfall” near 181 km (actually, 45 pixels of a 4-km length, or 180 km).
are produced. Although Fourier low-pass filtering of rainfall rate $R$ is, therefore, unacceptable, applying such a filter to log $R$ will at least avoid negative-filtered values for $R$. A truncated Haar wavelet approximation of the same rainfall field, omitting wavelet components for scales less than about 128 km $\times$ 128 km, is shown in Fig. A1c. This type of low-pass filter also smooths the sharp spatial gradients seen in Fig. A1a but localizes the distortions closer to the discontinuities. The relative disadvantage of a Fourier low-pass filter relative to a wavelet low-pass filter is determined by the magnitude of sharp gradients and discontinuities in the field to be studied.

The precipitation pattern in this case is not locally isotropic. In the northeast of the study area, it is stretched east–west. In the southeast, it is aligned from southwest to northeast. In this case, any attempt to apply appropriate anisotropic wavelets would need to adapt to the local precipitation pattern.

APPENDIX B

Wavelets and Filtering

A family of wavelet functions (a basis) is produced from a fundamental wavelet $\psi^0(x)$ by dilations and translations:

$$\psi^m(x - x') = 2^{-m/2} \psi^0 \left( \frac{x - x'}{2^m} \right). \quad (B1)$$

Admissibility conditions for choosing a fundamental wavelet are discussed by Mallat (1989). In this paper discrete power-of-two dilations are used, with a scale index $m$ that increases with increasing scale. For instance, a wavelet with scale index $m = 1$ is 2 times as wide as the fundamental wavelet, with a smaller amplitude. A wavelet with scale index $m = 2$ is 4 times as wide and one-half of the amplitude of the fundamental wavelet, and so on to the larger scales. By taking such discrete scales, the efficient wavelet transform algorithm of Mallat (1989) can be used.

Wavelet transform coefficients are given by

$$W(m, x) = \int_{-\infty}^{+\infty} f(x) \psi^m(x' - x) \, dx'. \quad (B2)$$

Each wavelet transform coefficient measures the similarity between a particular wavelet and the function $f(x)$ being studied. The wavelet is localized, and so the wavelet coefficient will depend on what $f(x)$ is like for that local range of position. The wavelet is also of a certain scale, so that the wavelet coefficient will describe the variability of $f(x)$ at similar scales. Where there are intense features of approximate length between $2^{m-1}$ and $2^m$ near $x$, the absolute value of $W(m, x)$ will be high. Where there is little variability at a particular scale and position, the absolute value of the associated wavelet coefficients is low. The associated inverse transform (reconstruction) formula is given by

$$f(x) = \sum_{m=0}^{\infty} 2^{-m} \int_{-\infty}^{+\infty} W(m, x') \psi^m(x - x') \, dx'. \quad (B3)$$

The integral gives the contribution of each scale number $m$ to the function. Each term in the summation acts as a wavelet bandpass filter of $f(x)$. In practice the summation is made for $0 < m \leq m_{\text{max}}$, because we cannot use scales larger than the length of measurements. The contribution of larger scales remains in a smoothed version of $f(x)$, which is like a running mean.

A wavelet power spectrum can be defined as

$$S(m) = \overline{W^2(m, x)}, \quad (B4)$$

where the overbar denotes a spatial average and $S(m)$ measures the importance of features with approximate length scales between $2^{m-1}$ and $2^m$. Similarly for two functions, say $f(x)$ and $g(x)$, one can define a wavelet cospectrum as

$$C_{fg}(m) = \overline{W_f^*(m, x) W_g(m, x)}, \quad (B5)$$

where the subscripts $f$ and $g$ denote the the two functions $f(x)$ and $g(x)$, respectively. Hudgins et al. (1993) showed that such a wavelet spectrum (or cospectrum) is equivalent to a Fourier spectrum (or cospectrum) smoothed by the Fourier spectrum of the wavelet function at each scale. The wavelet spectrum (or cospectrum) gives the contribution to the variance (or covariance) per logarithmic (roughly power of two) interval of scale.

Each fundamental wavelet $\psi^0(x)$ has an associated orthogonal smoothing function $\phi^0(x).$ As for the wavelets, a set of smoothing functions are produced by dilations and translations of $\psi^0(x).$ Greater mathematical detail is given by Burrus et al. (1998) and other similar texts. Convolution with the smoothing functions acts as a low-pass filter:

$$f^m(x) = \int_{-\infty}^{+\infty} f(x) \phi^m(x - x') \, dx'$$

$$= \sum_{m'=m+1}^{\infty} \overline{W(m', x') \psi^m(x - x') \, dx'}. \quad (B6)$$

The smoothed function $f^m(x)$ contains only those features represented by wavelet coefficients $W(m', \ldots)$, where $m' > m.$ The simplest admissible wavelet is the Haar function:

$$\psi^0(x) = \begin{cases} +1 & -\frac{1}{2} < x \leq 0 \\ -1 & 0 < x < +\frac{1}{2} \\ 0 & \text{otherwise} \end{cases} \quad (B7)$$

This can be used to represent a function as a sum of square wave pulses of different sizes at different posi-
The two-dimensional Haar wavelets and smoothing function are the square-wave pattern, and its smoothing function is a step function. The one-dimensional Haar wavelet is a single function, and one smoothing function are illustrated. Convolution of an image with the smoothing function of the two-dimensional Haar basis gives uniformly weighted block-average smoothing of an image for each scale.

Coefficients from the two-dimensional wavelet transform are used to calculate the two-dimensional wavelet spectrum and cospectrum:

\[ S(m) = 4^{-m} \sum_{i=1}^{2^m} [W(i, x, y)]^2, \quad \text{and} \quad (B10) \]

\[ \text{Co}_{ij}(m) = 4^{-m} \sum_{i=1}^{2^m} W(i, x, y)W'(j, m, x, y), \quad (B11) \]

where the overbar denotes a spatial (areal) average. The normalization factor becomes \(4^{-m}\) with the transition to the two-dimensional transform.

Section 2b derives optimal weighting factors \(w(m, T)\) for each scale \(m\) and a forecast lead time \(T\). This filtering can be implemented indirectly, avoiding explicit use of the inverse wavelet transform. Recall that as a result of the orthogonality of a wavelet and its smoothing function, the convolution of the image with this smoothing function acts as a low-pass filter:

\[
\begin{align*}
I^\gamma(x, y) &= \int_{-\infty}^{+\infty} I_f(x)\psi^\sigma(x - x', y - y') \, dx' \\
&= \sum_{m=0}^{\infty} 2^{-m} \int_{-\infty}^{+\infty} W(m', x')\psi^\sigma(x - x', y - y') \, dx', \\
&\quad \text{where } I^\gamma(x, y) \text{ is a smoothed version of the original forecast } I_f(x, y) \text{ with details of scales } m \text{ and smaller removed. A more intuitive illustration of this, with simpler notation, is shown in Fig. B2.}
\end{align*}
\]

The reconstruction formula can, therefore, be rewritten in terms of the uniformly weighted, block-averaged versions of the forecast image:

\[
I^\gamma(x, y, T) = \sum_{m=0}^{\infty} [w(m + 1, T) - w(m, T)]I_f(x, y, T). \quad (B13)
\]

Normally the scale weights at each scale will become larger with increasing scale index \(m\) at any time \(T\). Small scales will be less predictable and decorrelate faster than larger-scale features. In practice, there will be some maximum-scale index \(m_{\text{max}}\) beyond which wavelet coefficients are not calculated. For implementation with \(0 < m \leq m_{\text{max}}\), it is useful to set

\[
w(0, T) = 0, \quad w(m_{\text{max}}, T) = 1, \quad \text{and} \quad I^\gamma(x, y, T) = I_f. \quad (B14)
\]

The reconstruction formula is then restated as
In this way, the optimally filtered forecast is constructed as a weighted sum of the Haar wavelet smoothed images $I^m_f$ at each scale $m$.

$$I_f(x, y, T) = \sum_{m=0}^{\infty} [w(m + 1, T) - w(m, T)] I^m_f(x, y, T),$$  \hspace{1cm} (B15)$$

and the spatial smoothing function for the optimal filter is

$$\sum_{m=0}^{\infty} [w(m + 1, T) - w(m, T)] \Phi^m(x, y, T).$$  \hspace{1cm} (B16)$$

In this way, the optimally filtered forecast is constructed as a weighted sum of the Haar wavelet smoothed images $I^m_f$ at each scale $m$. 

![Schematic illustration of low-pass and bandpass filtering of the fast wavelet transform: S0 represents the image at finest resolution, S1 after smoothing to scale index $m = 1$, S2 after smoothing to scale index $m = 2$, and so on, and W1, W2, and W3 represent image features described by wavelet coefficients at scales $m = 1$, 2, 3, respectively.](image-url)