A New Multiscale Verification Method: The Error Spectra

Hsiao-ming Hsu and Mitchell W. Moncrieff
National Center for Atmospheric Research
PO Box 3000
Boulder, CO 80307

1. Introduction

Various statistical measures have been used to judge the forecast skill of a model. Many scores and skills can be found in the literature. Wilks (1995) and Jolliffe and Stephenson (2003) provide a general description on verification methods for forecasts. Based on temporal and/or spatial data from observations and models, these statistical measures usually indicate the overall quality of forecasts in terms of quantities difficult to measure directly (e.g., kinetic energy spectra) rather than variables such as precipitation, which can be verified using radar and rain gauge measurements. Our statistical approach is specifically focused on precipitation.

In our recent analysis of warm-season precipitation over North America (Hsu *et al.* 2006a), the temporal spectra for the latitudinally-averaged time series indicate a remarkable cross-scale self-similarity and periodicity, using both Fourier and wavelet transforms. Additional analysis of the spatial distributions of rainfall patterns with a 2-D FFT suggests a power-law scaling over a range of high wave numbers. Skamarock (2004) analyzed the kinetic energy of flight-track data and Weather Research and Forecast (WRF) forecasts, and showed similar power-law scaling in his data, using 1-D FFT.

For comparing numerical model results with observations, both spectral and wavelet analyses are effective tools to identify distinct features in spectra. The inverse wavelet transform has the advantage that it can further reconstruct the data at each frequency or wave-number effectively. The reconstructed data from observation and model can be quantitatively assessed at each frequency or wave-number using different statistical measures. In summary, our statistical-dynamical approach is a unique practical way to validate the space-time characteristics of precipitation in high-resolution NWP models.

Our new approach for scale-dependent verification was tested with the WRF experimental forecasts, which have already shown significant improvement compared to current operational forecasts. Therefore, we develop and use a multiscale verification method for precipitation distribution in the WRF forecasts. As an initial demonstration, and in view of the 1-month commitment, we perform only the time series analysis of surface precipitation. Measurements of precipitation from NEXRAD (Carbone *et al.* 2002) and rain gauges were used to evaluate the simulations.

2. Data

In this study, we use two observed rainfall data sets, one from *in situ* rain gauges and the other from remote-sensed radars. Model forecasts are predicted by the Weather Research and Forecast (WRF) model system.

During the BAMEX 2003 field campaign (Davis *et al.*, 2004; Done *et al.*, 2004), WRF provided experimental 36-hour forecasts every day from 00Z. The WRF forecasts used herein are the later 24-hour rain rates in the region where the 4-km resolution was configured (Fig. 1). The period we are interested is the complete one month during the BAMEX, June 2003. The forecasts retained are hourly WRF outputs. These hourly rainfall forecasts are interpolated onto rain gauge locations to form the time series for the temporal spectral analysis,.

Over the region which WRF forecasts were performed, there were rainfall measurements with the 1-minute resolution from a number of rain gauges at the Automated Surface Observing Systems (ASOS) locations. We impose a condition to retain high-quality data: If an individual rain gauge had data missing for more than 10% of its complete time series, it was excluded from the analysis. Thus, 93 gauges are retained, and their locations are plotted in Fig. 1. For these 93 time series, the missing parts of the time series are filled with zeros.

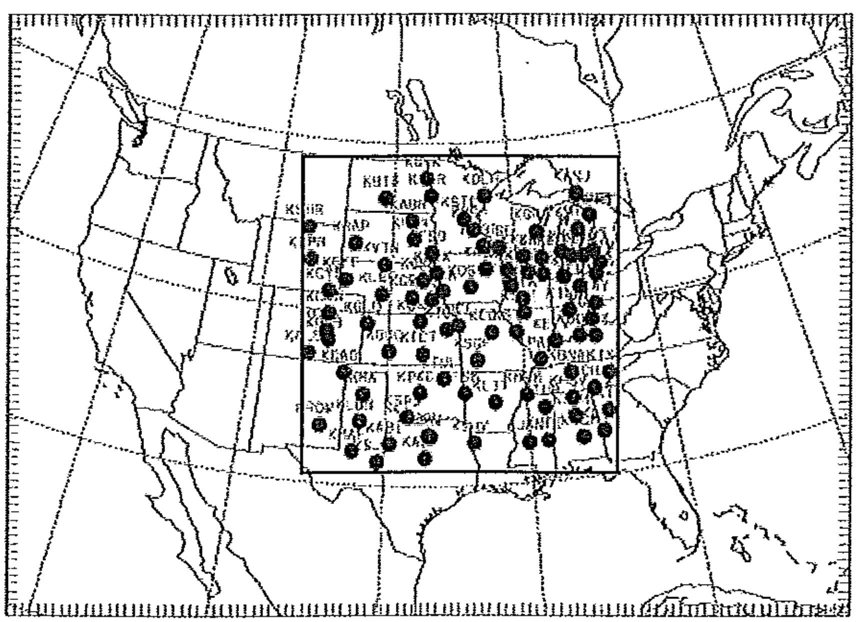


Figure 1. The 4-km WRF forecast region is indicated by the box, and the dots are the locations of the ASOS rain gauges.

Convective precipitation over the continental US have been measured widely with the NEXRAD radar network. Each individual radar observation has been interpolated into a national composite at a resolution of 0.02° in longitude and latitude (less than 2 km). The original composite radar reflectivity data from the NEXRAD is a WSI Corporation NOWrad product. These data have been used in the investigations of warm-season rainfall signals over North America by Carbone *et al.* (2000) to identify propagating convective streaks. Hsu *et al.* (2006a and 2006b) used the same data to investigate the temporal and spatial spectral structures, respectively. Hourly composite reflectivity data were interpolated onto rain gauge locations as done with the WRF forecasts. A Z-R relationship ($Z = 300R^{1.5}$) is applied to convert the reflectivity (Z in dbz) to the rain rate (R in mm/hr).

3. Temporal Spectra

After the rain-rate time series (WRF, ASOS, and NEXRAD) are prepared, they are decomposed using the continuous wavelet transform (CWT) with the Morlet wavelets as the mother function, as was done in Hsu *et al.* (2006a).

Since the WRF forecasts are available hourly, all the ASOS rain gauge and the NEXRAD radar data sets are coarse-grained into hourly rain rates. The analysis of the original ASOS and NEXRAD data will be described elsewhere.

The global temporal spectra are shown in Fig. 2. The ASOS rain gauge spectrum (solid curve) and the WRF spectrum (dashed curve) are similar for the low frequencies up to about 1 cycle/day (cpd), but have significant differences at the high frequencies. The original NEXRAD spectrum (light dotted curve) has the similar spectral structure with the ASOS one, but its variances are rather large. After the conversion from reflectivity to rain rate is altered ($Z = 500R^{1.5}$). The new NEXRAD spectrum (dotted curve) is very close to the ASOS one. Therefore, the measured spectra (either observed or predicted) are consistent, while the WRF spectrum has significant departures at high frequencies.

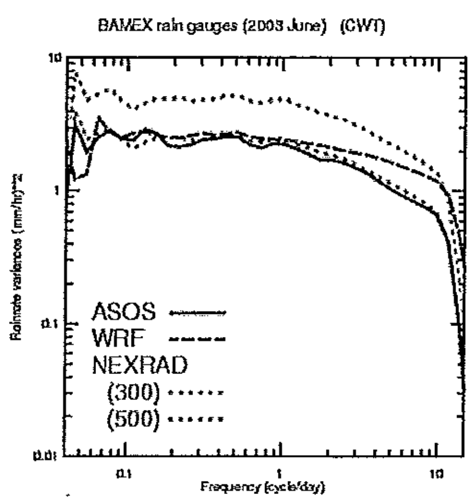
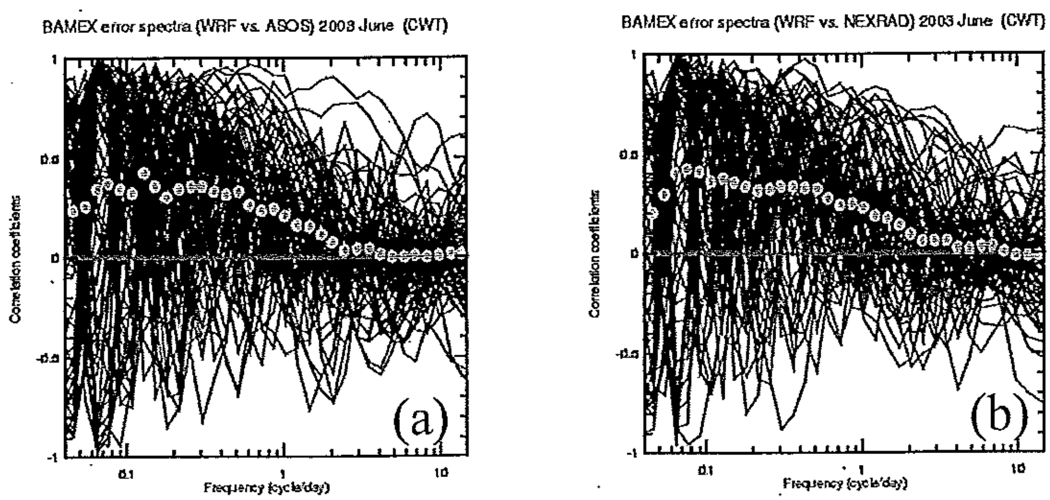


Figure 2. The temporal spectra of rain-rate variances for ASOS rain gauges (solid curve), NEXRAD radars (dotted curves), and WRF forecasts (dashed curve).

4. Error Spectra

Although the similarities and differences between model forecast and observation are identified in frequency space, the relative error measure of model forecasts in each frequency should shed further light than just one single error measure for the whole data set. Here is what we propose. At each rain gauge, each time series (either observed or predicted) is decomposed into frequency space using CWT. At each frequency, the spectral coefficients are inversely transformed back into the physical space to form the time series for that particularly frequency. This procedure can be repeated for each frequency of the spectrum. Then at each corresponding frequency, the statistical measure between the forecasted and the observed is calculated to produce the *error spectrum*: by definition, a frequency dependent statistical measure. In other words, the error spectrum shows the temporal scale-dependence of the statistical measure. Here, we use the correlation coefficient (CC) as an example for the statistical measure.



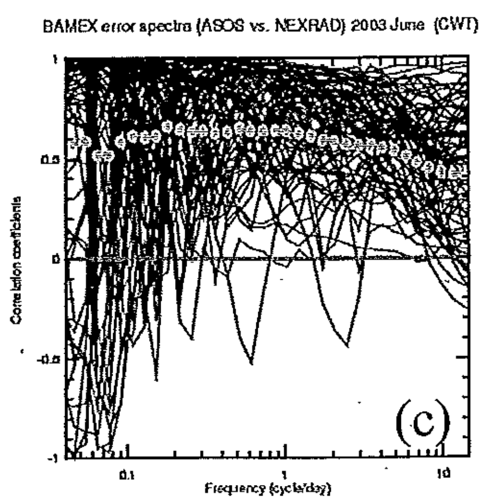


Figure 3. Error (correlation coefficient, CC) spectra of rain rates: (a) CC spectra between WRF forecasts and ASOS rain gauges, (b) CC spectra between WRF forecasts and NEXRAD radars, and (c) CC spectra between ASOS rain gauges and NEXRAD radars.

First, we compare the WRF forecasts with the ASOS rain gauge measurements. At each of the 93 ASOS sites, each error spectrum is plotted in Fig. 3a. The dots depict the averaged error spectrum. Obviously, for the frequencies lower than 1 cpd, the CC varies between 0.3 and 0.5. However, there is essentially no correlation for frequencies higher than 1 cpd. The CC spectra in Fig. 3b display the comparison between the WRF forecasts and NEXRAD measurements. They are similar to those resulted from the WRF-ASOS comparison. How do ASOS and NEXRAD date sets compare? The CC spectra illustrated in Fig. 3c indicate almost evenly distributed CC across frequencies at a level of around 0.5. These results confirm the conclusion from the temporal spectra that the forecasts are more skillful at low frequencies than at high frequencies. The error spectra provide an additional quantitative measure in regard to temporal scales.

5. Discussion

The idea of the error spectra can be expanded to two dimensions to investigate the statistical measures in the spatial space in terms of horizontal wavenumbers. For the rainfall patterns during June 2003, two power-law scaling are found with a spectral break at 20 km, and their exponents are -2/3 and -4/3 for the low and high wavenumbers, respectively (Hsu *et al.,* 2006b). The next step is to perform the spatial error-spectrum analysis.

The generally low correlation coefficients (~ 0.5) between ASOS and NEXRAD data in the average quantify the difficulty in comparing rainfall measurements.

It is well known that the point-to-point evaluation of model forecasts is difficult for various reasons, particularly for rainfall verification, because of the intermittency of the rainfall events. Comparison of temperature and wind speed between model forecasts and measurements would be interesting because they are much less intermittent.

REFERENCE:

Carbone, R., J. D. Tuttle, D. A. Ahijevych, and S. B. Trier, 2002: Inferences of predictability associated with warm season precipitation episodes. *J. Atmos. Sci.,* **59,** 2033–2056.

Davis, C., N. Atkins, D. Bartels, L. Bosart, M. Coniglio, G. Bryan, Cotton W, Dowell D, Jewett B, Johns R, Jorgensen D, Knievel J, Knupp K, Lee W-C, McFarquhar G, Moore J, Przybylinski R, Rauber R, Smull B, Trapp J, Trier S, Wakimoto R, Weisman M, Ziegler C., 2004: The bow-echo and MCV experiment: Observations and opportunities. *Bull. Amer. Meteor Soc.,* **85**, 1075-1093.

Done, J., C.A. Davis, M. Weisman, 2004: The next generation of NWP: Explicit forecasts of convection using the weather research and forecasting (WRF) model. *Atmos. Sci. Let.* **5**: 110–117.

Hsu, H.-m., M.W. Moncrieff, W.-w. Tung, and C. Liu, 2006a: Temporal Variability of Warm-Season Precipitation over North America: Statistical Analysis of Radar Measurement. *J. Atmos. Sci.* (in press).

Hsu, H.-m., M.W. Moncrieff, P. Sullivan, W.-w. Tung, M.J. Dixon, and J.D. Tuttle, 2006b: Spatial Spectral Structure of Warm-Season Rainfall over North America. *J. Atmos. Sci.* (in preparation).

Jolliffe, I.T. and D.B. Stephenson, 2003: *Forecast Verification: A Practitioner's Guide in Atmospheric Science*. John Wiley, 240*pp*.

Skamarock, W.C., 2004: Evaluating mesoscale NWP models using kinetic energy spectra. *Mon. Wea. Rev.,* **132**, 3019-3032.

Wilks, D.S., 1995: *Statistical Methods in the Atmospheric Sciences*. Academic Press, 467*pp*.