Selection technique for thinning satellite data for numerical weather prediction

Christian Alcala, Ross N. Hoffman, and S. Mark Leidner
Atmospheric and Environmental Research, Inc.
131 Hartwell Avenue, Lexington, MA 02421

1. Summary

Operational weather prediction centers use only a fraction of observations of the atmosphere and the earth's surface that are made by satellite, *in situ*, and ground-based instruments. We are investigating the use of wavelet analysis to develop an adaptable selection method based on the local information content in a satellite data scene to determine the density of observations to use. This investigation supports and enhances Earth science capability by:

1. Improving the selection and impact of the vast, information-rich and valuable satellite observations of the Earth system.
2. Combining mature technologies (atmospheric data assimilation and wavelet analysis) for a novel and practical use.
3. Raising the technology readiness level (TRL) of this technique to a working prototype in a realistic setting.

Current practice at operational weather centers reduces today's enormous volume of satellite data to practical levels by regular decimation, i.e., keeping every 2nd or 4th or 6th point along and across the satellite's data swath. Our work investigates an adaptable approach to select data. We are using the continuous wavelet transform (CWT) [1] in two dimensions (*x,y*) to identify features of interest for more informed data thinning. Retrieved wind speeds from NASA's SeaWinds scatterometer on QuikSCAT [2] provide the input data for the thinning technique in our study. The amplitudes of the wavelet coefficients from two passes of the CWT (top-to-bottom and left-to-right) are summed to identify edges and gradients in the satellite-observed wind speed field. The CWT provides information on six spatial scales (25, 50, 100, 200, 400, 800 km), so features of interest can be identified on each of these scales. Once features are identified at each spatial scale, the satellite data are decimated to a density appropriate to the associated spatial scale. The final data selection is the union of all points selected at every scale. We test our thinning technique by data assimilation in atmospheric models with a 2d-variational method [3]. The baseline case assimilates all available data (ALL). This provides a "best" analysis since it uses all available data. But it is also computationally expensive. Experiments assimilating thinned subsets of the data by regular decimation and wavelet-based selection are evaluated for information content. Our goal is to retain as much information as possible in the data assimilation analysis but by only using 3-5% of the data through wavelet-based selection.

Our results to date show that wavelet-based selection is roughly equivalent to regular decimation to every 6th or 8th datum. The smallest scale wavelet information (25 and 50

km) is ignored in results to date because the signal-to-noise ratio is small. A new method to remove spurious wavelet transform modulus maxima (WTMM) will allow the use of features at these scales and is expected to achieve better results.

2. Wavelet power spectrum analysis applied to satellite wind speeds

Our data-thinning algorithm uses the CWT to obtain a measure of the local information content of the data. This approach was first developed in the context of edge detection [4]. In particular, the CWT has several important advantages over the more commonly used discrete wavelet transform (DWT). Both transforms provide a complete and invertible representation of the data. However, while the DWT uses a set of orthogonal wavelet bases to obtain the most compact representation of the image, useful for data compression, the CWT uses a set of nonorthogonal wavelet frames to provide a highly redundant representation. The consequence of this redundancy is that the CWT gives a wavelet coefficient at each analysis scale for each pixel in the image, allowing us to characterize the local information content. In addition, this redundancy improved the stability of the reconstruction (inverse transform) in the presence of noise. Using a wavelet based on the conventional Canny edge detector allows us to simultaneously detect, localize, and characterize the edges in the observation data. Following the approach described by Mallat [1], we first apply the CWT using a wavelet oriented across the satellite track and then repeat using a wavelet oriented along the satellite track. We then take the coefficients from these two-orthogonal directions to produce a wavelet amplitude and phase for each pixel at each scale.

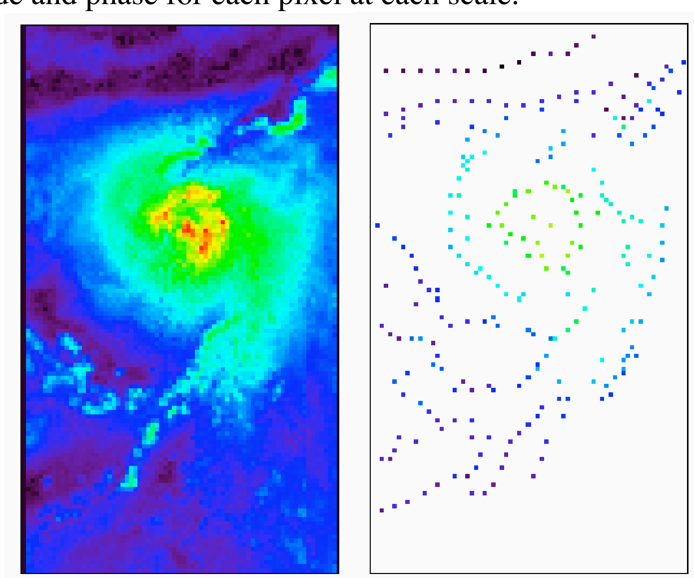

Figure **1**. SeaWinds wind speeds (left) and selected wind speeds (right) for Typhoon Meranti at 0600 UTC 7 August 2005. Wind speeds from 0 to 30 m/s are shown as colors from purple to red.

After analyzing the data with the CWT, we have developed the following technique to identify the features of interest and select the data points along these features:
1. Identify the wavelet transform modulus maxima (WTMM) for each scale.
2. Use noise thresholding to remove spurious WTMM from noise.
3. Chain neighboring WTMM at each scale to produce ridges.

4. Compare co-located ridges at different scales to characterize structures.
5. Select data locations along WTMM ridges.

Because noise introduces multiple small-scale edges within the data, we need to develop a technique for distinguishing between WTMM produced by signal and WTMM produced by noise at these scales. Simple thresholding will not work since both types of WTMM are of equal magnitude. However, the WTMM associated with actual geophysical structure will track down from larger to smaller scales. Therefore, we use an approach similar to the scale multiplication method of Zhang and Bao [5] to partition the signal from the noise at these small scale sizes.

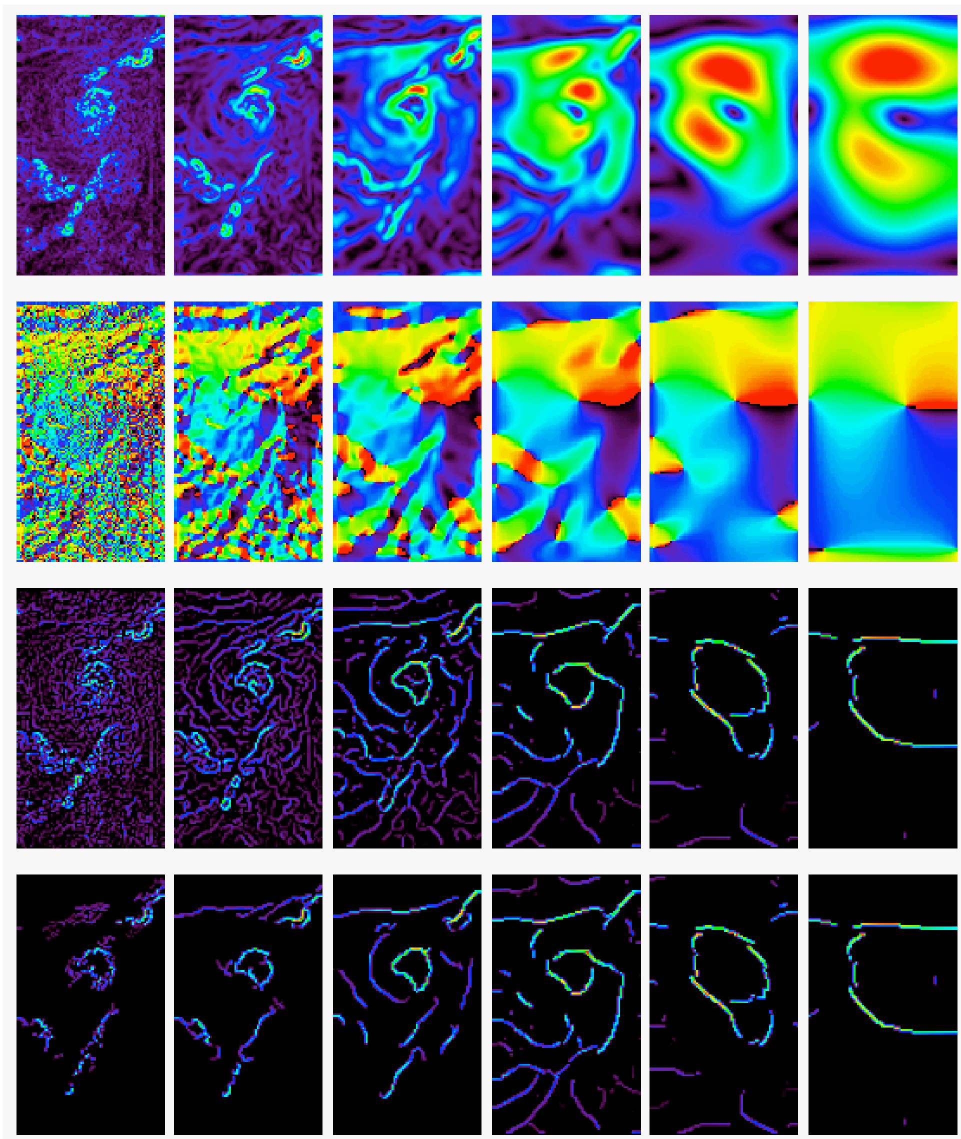

Figure **2**. Two-dimensional wavelet analysis of Typhoon Meranti ocean surface wind speeds (from Fig. 1). See text for detailed description.

In Figs. 1 and 2, we apply our data-thinning algorithm to an example of satellite-observed ocean-surface wind speeds obtained during Typhoon Meranti, 1800 UTC 7 August, 2004. Figure 1 shows the satellite image in the left panel and the results of our analysis in the right panel. The number of data values selected by the algorithm for this example corresponds to about 2.5% of the total number of data points. Figure 2 shows the intermediate results generated by the wavelet analysis for the image in Fig. 1. The top two rows in Fig. 2 plot the magnitude and phase of the two-pass CWT operation for the scales of 25, 50, 100, 200, 400, and 800 km. The bottom two rows show the results of the data-thinning algorithm before data location selection both without and with the additional noise-partitioning step for all of the scales. By comparing these two rows we see a significant reduction in number of WTMM ridges in the smallest three scales (25, 50, and 100 km). We are presently in the process of integrating this noise-partitioning step into the data-thinning algorithm for comparison with the currently used decimation in the data assimilation analysis.

3. Analysis of thinned and decimated SWS data

We conducted a search for appropriate cases for data assimilation and forecast experiments. We want to isolate the impacts of data thinning from other contaminating influences as much as possible, so we have adopted a selection strategy that only considers satellite data within ~10 minutes of synoptic times (i.e., 00, 06, 12, and 18 UTC). This minimizes the time difference between satellite observations and available global analyses of the surface wind field so that meteorological features of interest are aligned as closely as possible. We selected four cases that represent a wide range of meteorological conditions and have sufficient observations at subsequent times for validating the mesoscale forecasts. One of the cases has already been presented in Fig. 1.

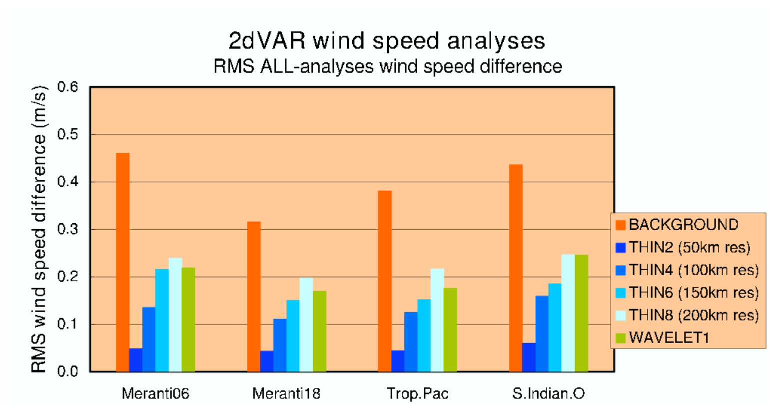

Figure **3**. RMS difference between ALL 2dVAR analysis and analyses produced using various data thinning strategies for four case studies.

The design of experiments for 2dVAR analyses includes four cases, six thinning approaches and two assimilation experiments. The cases include Typhoon Meranti at 0600 and 1800 UTC 07 August 2004, an anticyclone in the South Indian Ocean at 1200 UTC 31 October 2004, and a very light wind case in the tropical Pacific at 0000 UTC 5 March 2005. Note that the WAVELET1 experiments retain about 3.5% of the data (the union of THIN8 and wavelet-selected data). Taking the analyses that use all available data as a "best-performance" benchmark (analysis ALL), the 2dVAR analyses from wavelet-selected data (WAVELET1) have accuracy (in an RMS sense) that falls between analyses produced by decimating the data to every 6th and 8th data point (THIN6 and THIN8 vs. WAVELET1 in Fig. 3). We believe that the new noise thresholding described in the previous section will allow smaller scale features to be established in the analyses that will improve the RMS fit to the observations.

REFERENCES

[1] Mallat, S., 1998: *A wavelet tour of signal processing*, 2nd edition, Academic Press.

[2] Spencer, M. W., C. Wu, and D. G. Long, 2000: Improved resolution backscatter measurements with the SeaWinds pencil-beam scatterometer. *IEEE Trans. Geosci. Remote Sens.*, **38**, 89–104.

[3] Hoffman, R. N., S. M. Leidner, J. M. Henderson, R. Atlas, J. V. Ardizzone, and S. C. Bloom, 2003: A two-dimensional variational analysis method for NSCAT ambiguity removal: Methodology, sensitivity, and tuning. *J. Atmos. Oceanic Technol.*, **20**, 585–605.

[4] Mallat, S. and S. Zhong, 1992: Characterization of signals from multiscale edges, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **14**, 710-732.

[5] Zhang, L. and P. Bao, 2002: Edge detection by scale multiplication in the wavelet domain, *Pattern Recognition Letters*, **23**, 1771-1784.