12.1

A Real-Time System to Estimate Weather Conditions at High Resolution

Peter P. Neilley[1]
Weather Services International, Inc.
Andover, MA 01810

And

Bruce L. Rose
The Weather Channel
Atlanta, GA

1. Introduction[1]

The purpose of this paper is to describe an operational system used to estimate current weather conditions at arbitrary places in real-time. The system, known as High Resolution Assimilation of Data (or HiRAD), is designed to generate synthetic weather observations in a manner equivalent in scope, timeliness and quality to a arbitrarily dense physical observing network. Our approach is, first, to collect information from a variety of relevant sources including gridded analyses, traditional surface weather reports, radar, satellite and lightning observations. Then we continuously synthesize these data into weather condition estimates at prescribed locations. An operational system based on this approach has been built and is commercially deployed in the United States.

In most regards, our approach is analogous to modern data assimilation techniques. However, since our objective is to produce consumer-oriented surface weather information, our requirements differ from that of standard assimilation systems (e.g. objective analysis, variational assimilation, Ensemble Kalman filter) in the following ways:

a) We do not create a full 3-dimensional estimate of the state of the atmosphere; we instead focus on the observable weather at the earth's surface (the so-called current conditions).

b) We do not necessarily produce weather observations on a regular grid, but at an irregular set of arbitrary locations or points that are relevant to the consumers of the information.

c) In addition to producing quantitative observational elements (e.g. temperature, pressure and wind speed) our system produces common, descriptive terminology of the sensible weather such as "Thundershowers", "Patchy Fog", and "Snow Flurries".

d) We do not strive to produce a state of the atmosphere optimized for fidelity with Numerical Weather Prediction (NWP) models. Instead, the system is optimized to produce the most accurate estimate of the observed state at the surface that can be used directly as consumer weather information.

Section 2 describes the methodology used to estimate the weather conditions. Section 3 provides some examples. Section 4 reports on analyses to ascertain the quality of the observations. Section 5 is a summary and comments on our future work efforts. Complementary papers (Rose et al, 2006; Koval et al., 2006) on HiRAD are found elsewhere in these conference proceedings.

[1] Corresponding author and presenter. Peter P. Neilley, Weather Services International, 400 Minuteman Rd, Andover, MA 01810, USA; e-mail: pneilley@wsi.com

2. Methodology

The basic methodology employed in HiRAD is that of a layered approach in which a sequence of refinements to a first-guess is made using a set of relevant data sources.

2.1 *First-guess values*. The initial values used by HiRAD are extracted or derived from the RUC 13 km[2] 0-3 hr forecast data. These data are collected hourly from the U.S. National Weather Service (NWS) ftp servers and are typically available about :45 minutes past each hour. The RUC output grids from the hours surrounding the valid time of the synthesized observation are linearly interpolated in time. In most circumstances, this interpolation is done between either the 0-hr (analysis) and 1-hr forecast, or between the 1- and 2-hr forecasts.

Table 1: Quantitative Variables Estimated by HiRAD
Temperature
Dew Point Temperature
Wind Speed
Cloud Cover (Percent of Sky)
Probability of Precipitation Occurring
Probability of Thunder Heard
Conditional Probability of Rain if Precipitation is Occurring
Conditional Probability of Snow if Precipitation is Occurring
Conditional Probability of Ice, Freezing Rain or Mixed Precipitation if Precipitation is Occurring
Probability of Fog
Precipitation Rate
Visibility

[2] The data used is based on the RUC 13-km resolution data. However, as of this writing, only the degraded 20-km version of the output grids is published by the NWS.

Table 1 lists the quantitative observation variables that are extracted from the RUC. These variables are used in subsequent processing by HiRAD and are the basis for deriving the final output of HiRAD

2.1 *Spatial Downscaling*. The RUC first-guess values are spatially refined using statistical downscaling based on high-resolution terrain and climatological information. This downscaling partly accounts for variance in the weather due to topographical gradients not captured in the RUC grids. This downscaling is most relevant in rapidly varying terrain such as the Intermountain Western United States. The high-resolution climatological database used is similar to that described by Daly (1994).

2.2 *Correction using recent observations*. The next step in the process involves correcting the downscaled first-guess values based on recent, nearby surface observations. For each recent observation, a difference between the refined first-guess values and the actual observation is computed. These differences are spatially interpolated using a Kriging technique (c.f. Oliver and Webster, 1990) applied at the locations of the synthetic observations. The interpolated correction values are then added as adjustments to the refined first-guess values at this target point. Observational covariances needed for the Kriging interpolation are assumed to vary inversely proportional to the distance between locations. In future versions of the system more realistic covariance matrices may be used. The sites included in the interpolation are determined using an iterative process that selects additional observation sites that exhibit maximum covariance with the interpolation site (most relevant) but minimum covariance with already selected sites (most unique). Fig. 1 illustrates an example of the site selection process.

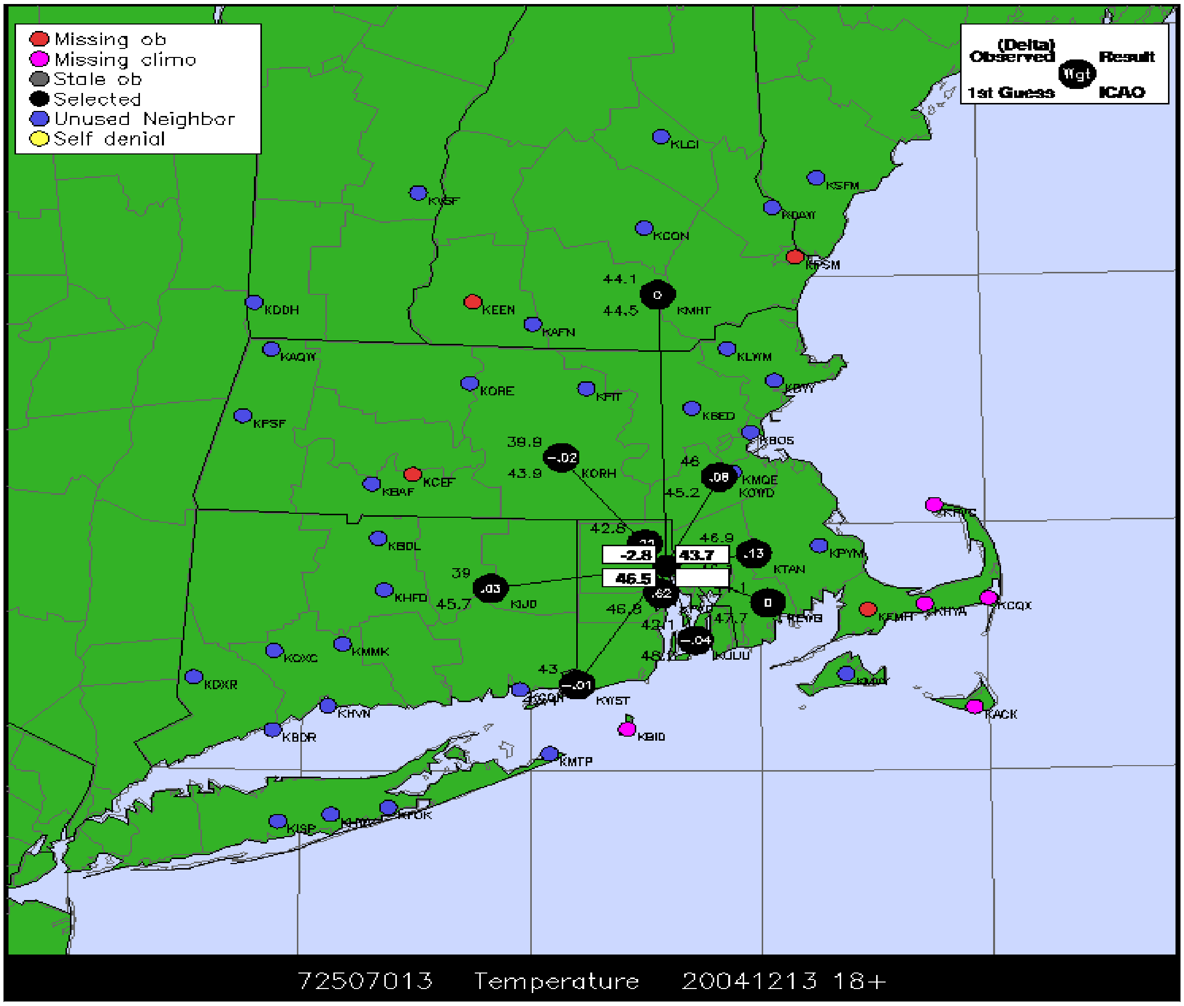

Fig 1. Example of the site selection process outcome in the Observation Engine. The example shows the sites considered for interpolation (all dots) and the sights actually chosen (black dotes with Kriging weights in the middle). Other sites not choose are color coded for the reason they were not included.

The Kriging technique assures that synthesized observations created at points collocated with actual observations will be equivalent to the actual observation (at least through this phase in the process). Achieving this was considered an important characteristic of HiRAD. We note that this is typically not the case with most data assimilation systems as those systems consider the error statistics of the observation, the error statistics of the background field and the meteorological consistency of the observation when incorporating the observation. Hence, unlike HiRAD, most data assimilation techniques generally do not replicate an observation even if that observation is collocated with an assimilation data point.

The current version of HiRAD uses only surface observations in METAR form for this error correction step. The METAR observations are subject to quality control (QC) before being used. This QC is based on differences between the observations, the downscaled RUC first-guess values and statistics derived from histories of these differences. Observations that fall outside configurable ranges of variance from the downscaled first-guess are rejected.

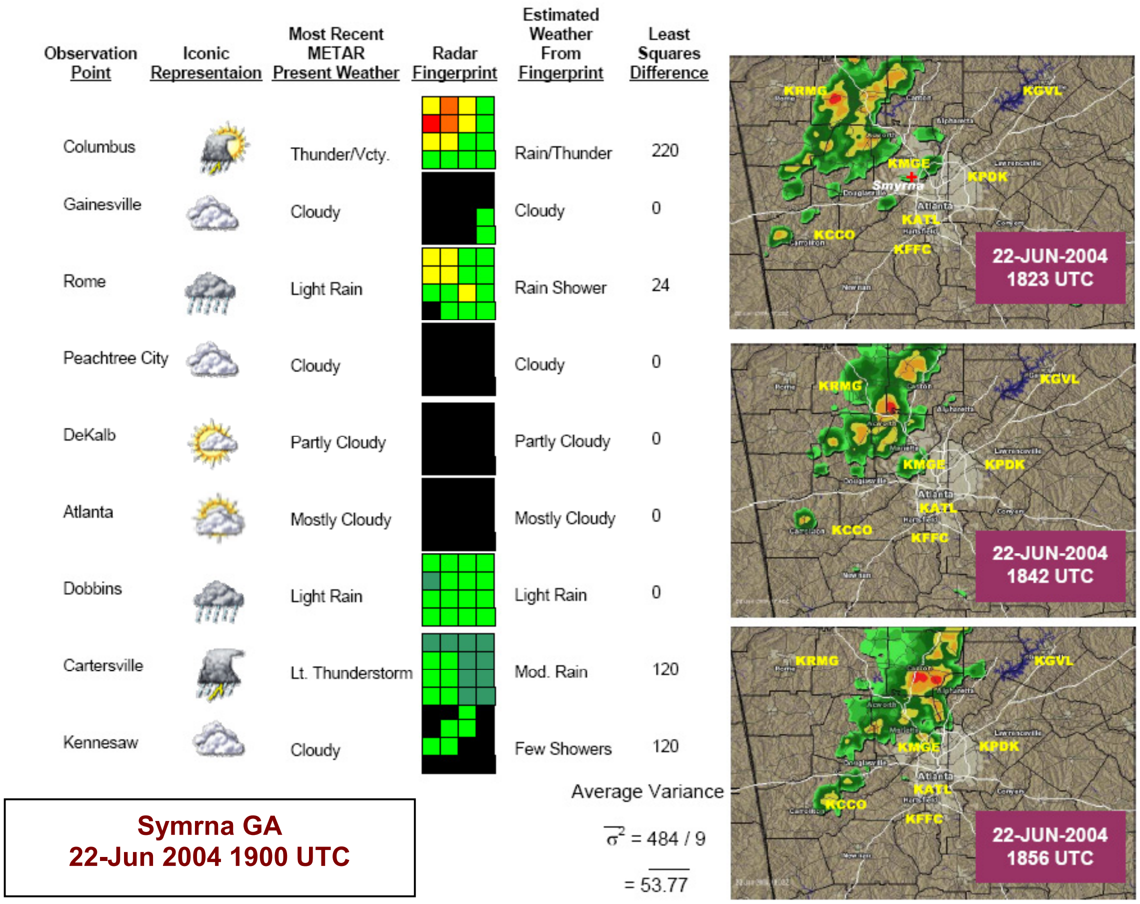

Fig 2. Illustration of the fingerprinting and calibration processes of the Observation Engine for a case near Smyrna GA on 22 June 2004, 1900 UTC. The images on the right show the time series of radar data leading up to the observation time. For each neighboring site near Smyrna, the columns on the left indicate METAR weather observations, local radar snapshots, fingerprinter-estimated weather and the "distance" between the two (see text).

Before inclusion in HiRAD, the METAR cloud reports are augmented with recent mid- and upper-level cloudiness estimates derived from GOES satellite observations. These satellite enhancements correct for the deficiency in fully-automated ASOS-based METAR observations in the United States that are unable to detect cloud amounts or ceilings above 12,000 feet AGL. This cloud augmentation step is the only direct use of satellite data in the current version of HiRAD. However, satellite data is used extensively in the initial RUC-based grids.

2.3 *Lighting Data*. Current lightning ground strike data (and a proportion of the cloud-cloud and in-cloud) from the United States Precision Lightning Network (USPLN) operated by Weather Decision Technologies of Norman, Oklahoma (WDT) is then incorporated in the system. For each observation site, a probability of thunder is computed based on the number of nearby, recent lightning strikes recorded. Lightning strikes within 30 nautical miles of a synthesized observation location are considered.

2.4 *Radar Data Overlay*. The inclusion of current weather radar data in the observation estimation process is one of the most important and complex components of HiRAD. Together, radar and lightning data generally represent the most contemporary data used by HiRAD and hence often contribute significantly to the observation estimation process. In the present version of HiRAD, a 2-km, 5-minute, quality-controlled mosaic of the U.S. WSR-88D radars is used. This mosaic includes estimates of precipitation type based on nearby surface weather observations. The mosaic is produced and commercially available from WSI Corporation of Andover, Massachusetts.

There are three basic steps in the inclusion of the radar data in HiRAD. These are fingerprinting, calibration and integration.

2.5.1: Fingerprinting: The purpose of the radar fingerprinter is to estimate the general weather conditions in the vicinity of the observation site based on the character of the nearby radar data and ancillary information from the first-guess quantities. The radar fingerprinter produces a numeric value representing a description of the estimated weather conditions using a complex, proprietary scheme beyond the scope of this paper. The potential outcomes of the radar fingerprinter is quite broad and can include “Partly Cloudy”, “Few Showers”, “Rain with Thunder”, “Heavy Thunderstorm”, “Snow Flurries”, etc.

2.5.2: Calibration: The second step is a calibration of the radar’s ability to accurately

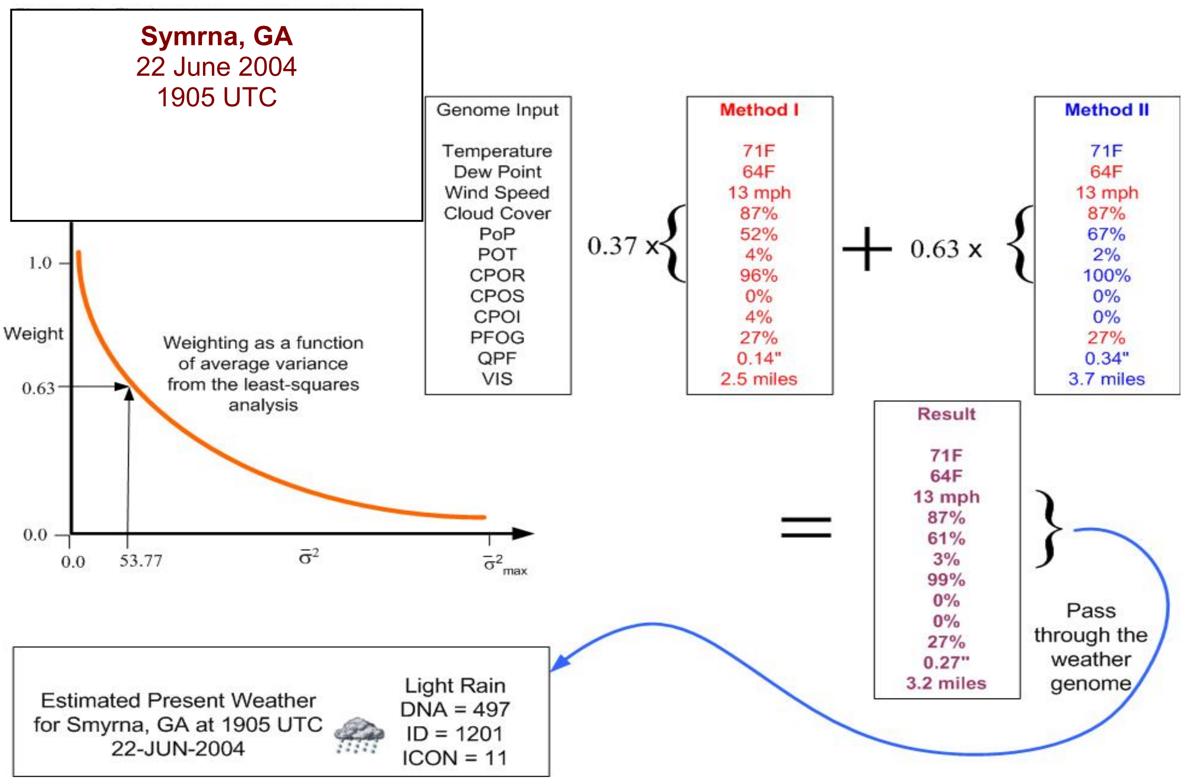

Fig 3. Illustration of the radar fingerprinting integration process for the example shown in Fig. 2. The figure illustrates the weighting function used to combine the observation estimates, the individual vector of observation quantities, the combined vector and the resulting end result

estimate the weather. This is accomplished by having the radar fingerprinter estimate the weather conditions at the time and location of each METAR weather observation in the (time and space) vicinity of the synthetic observing site. A multi-parameter “distance” between the fingerprinter-estimated and actual weather is computed at each site and a variance across all such relevant distances is determined. The parameters used to compute the distance are non-dimensionalized versions of the values listed in Table 1. Fig 2. shows an example of the fingerprinting and calibration process for a scattered thundershower case near Symrna, GA.

2.5.3 Integration: The purpose of the calibration step is to ascertain whether the actual observed weather and that estimated from the radar data are consistent. The result of this assessment is continuous, and hence a conclusion on the trustworthiness of the radar data is not always obvious. Hence, we have adopted a fuzzy-logic approach in combining the weather estimates from the radar-fingerprinter and the observation estimate before the fingerprinter. A weighted-average of these two estimates is computed, with the weight

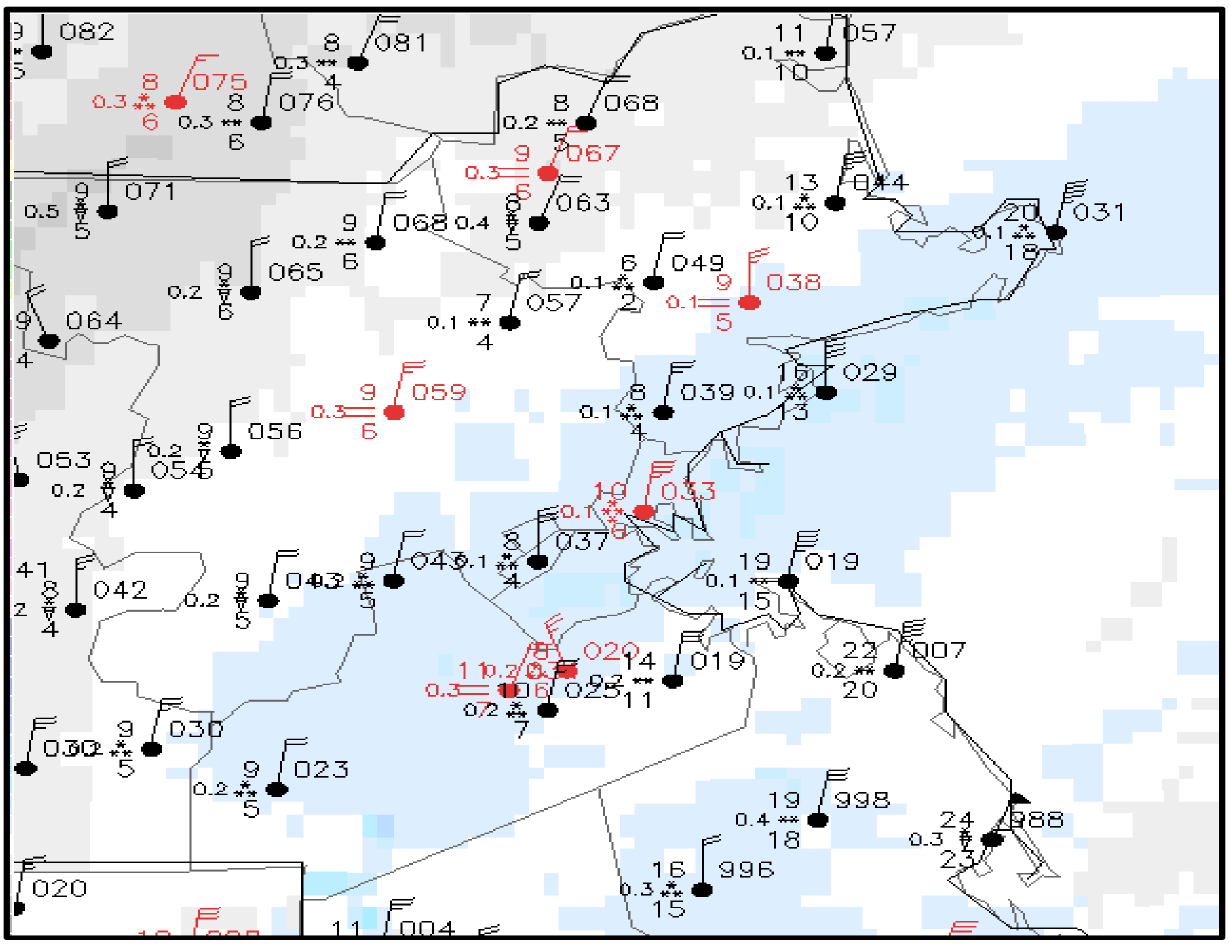

Fig 4. Example output from HiRAD over eastern Massachusetts on 23 January 2005 at 1400 UTC. Sites plotted in red represent actual observations from ASOS/METAR sites, while black sites represent estimated weather observations. Data is plotted using standard meteorological conventions, with values shown in English units (F, knots, miles). The background field represents the radar reflectivity with increasing values from gray to blue.

of the radar-data being proportional to the reciprocal of the variance computed during the calibration step. Fig. 3 illustrates this process for the case shown in Fig. 2.

2.5 *Final Processing*. A number of final steps in the observation estimation process are then undertaken. This includes estimation of the final weather description determined from the final values of the basic quantitative variables, several consistency checks on these values, and an "aberrant" weather override.

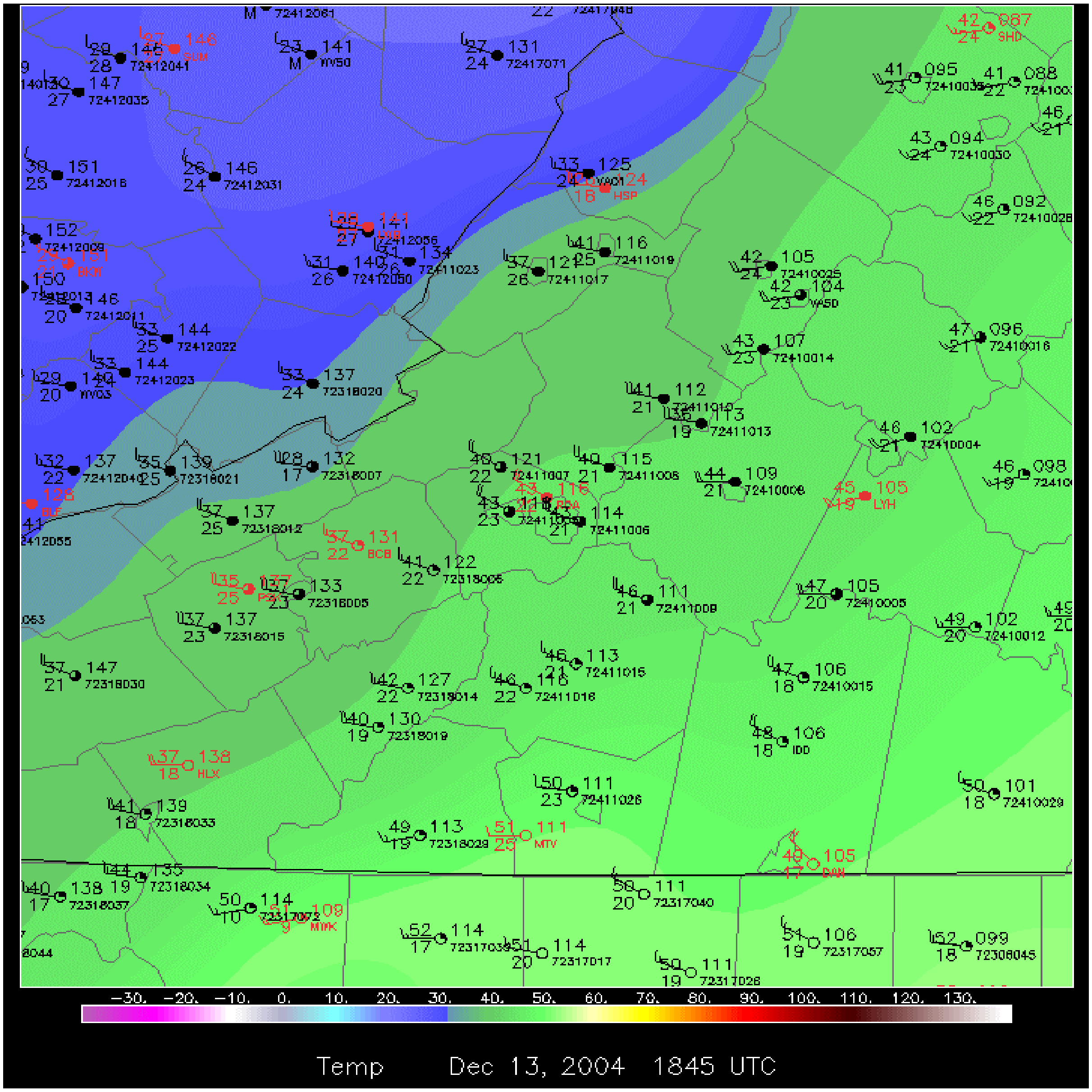

Fig. 5. As in Fig 4, but for a dry cold front passage over the mountainous portions of western Virginia on 13 December 2004 at 1845 UTC. In this case, the observations plotted in red are from sites that make actual weather observations, but are estimated by the observation at this particular time. Color filled contours underneath the station plot are for temperature. Considerable variance in temperature data is evident associated with the cold front passage and the terrain changes in the region.

This latter step attempts to account for types of the weather that can occur but cannot be adequately described by the current processes. Examples include smoke and blowing dust. The aberrant weather override checks nearby, actual-observation sites for repeated reports of certain types of weather, if these special weather types are found to be widespread or pervasive, they incorporated into the final weather descriptor of the synthetic point.

Lastly, the final estimated observational data are formatted and packaged for output.

3. Examples.

Fig. 4 is an example of HiRAD output during the heart of an especially intense winter snowstorm that struck Eastern Massachusetts this past winter season. The data is shown using standard meteorological plotting conventions. The radar data at the time is plotted underneath. The sites plotted in red represent actual weather observations from the ASOS/METAR network, while sites in black represent synthetic conditions from HiRAD. We note that many of the ASOS sites were incorrectly reporting the weather as fog or haze during this time as the intense, heavy, blowing snow caused instrumentation and possible algorithm processing problems at the automated ASOS sites. However, sufficient influence from the radar fingerprinting and calibration process enabled all of the synthetic sites to report significant snowfall rates. Note, that the synthetic weather observations contain variance that cannot be explained by simple interpolation of the actual observations. This is most evident near the coast, where there is a

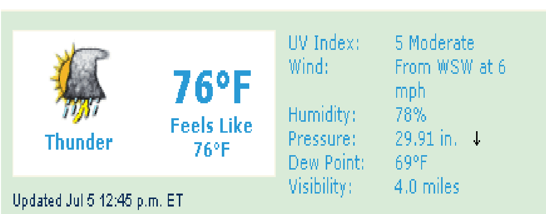

Fig 6. Example of a textual output of the Observation Engine for Mount Vernon, OH on 5 July 2005, 1745 UTC during the passage of a thunderstorm complex.

significant increase in wind speed and temperature, consistent with the expected weather in such circumstances. Additional variance in the intensity of the precipitation is evident as well and consistent with the radar data from this event.

Fig. 5 shows another example during a dry passage of a cold front. Variations evident in this example are primarily due to the cold front and terrain of the region. Finally, Fig 6 shows an example of a textual representation of HiRAD output available on a commercial web site.

4. Skill of HiRAD Estimates

A direct, quantitative skill estimate of HiRAD is somewhat complex since an independent validation dataset is generally not available. Existing surface weather observations already are included in HiRAD and hence comparisons with these observations are not especially meaningful. Comparisons with secondary observation networks or mesonets was considered but not completed due to the uncertainty in the quality of these data. Furthermore, most mesonet sites do not report the set of weather parameters we are interested in such precipitation or present-weather type, cloud cover, and visibility.

Therefore, to quantitatively assess the quality of HiRAD, a version of the system was developed that produced synthesized observations at locations reporting actual METAR observations, but without using the observations from those sites. Comparisons could then be made between the estimated observation and the unused actual observations. These data denial experiments were implemented side-by-side with the actual HiRAD system.

Results from these data denial experiments for December 2004 have been generated and aggregated over approximately 200 sites in the U.S. For temperature, the mean difference between the estimated and actual observation was found to be 0.99 C. The average was slightly higher in mountainous regions. We note that the stated accuracy of the weather observations from the U.S. ASOS system is 0.6 C. This suggests that HiRAD is producing temperature estimates not much less accurate than actual ASOS observations.

Subsequent use of the data denial system has focused on analysis of cases with large, significant differences between the estimated and actual observations. In nearly all cases, the source of the difference can be attributed to an error in (reporting of) the observation. More complete verification results can be found in the companion paper by Koval et al. (2006).

5. Future Work.

In the next phase of HiRAD development the point-based system will be transformed to a high-resolution grid (2.5 km covering the CONUS) of some 1.9 million grid point intersections. We have chosen to derive this grid explicitly. That is, the core interpolation and radar fingerprinting algorithms will produce synthetic observations at each grid point as opposed to a simple spreading or downscaling of the point-based data. This is much more arduous but we believe strongly it will better capture small-scale features or gradients in the sensible weather, wind flow, and temperature. Early tests show this entire grid can be produced and disseminated on the order of five minutes elapsed time.

In this phase III HiRAD development more attention will be placed on precipitation

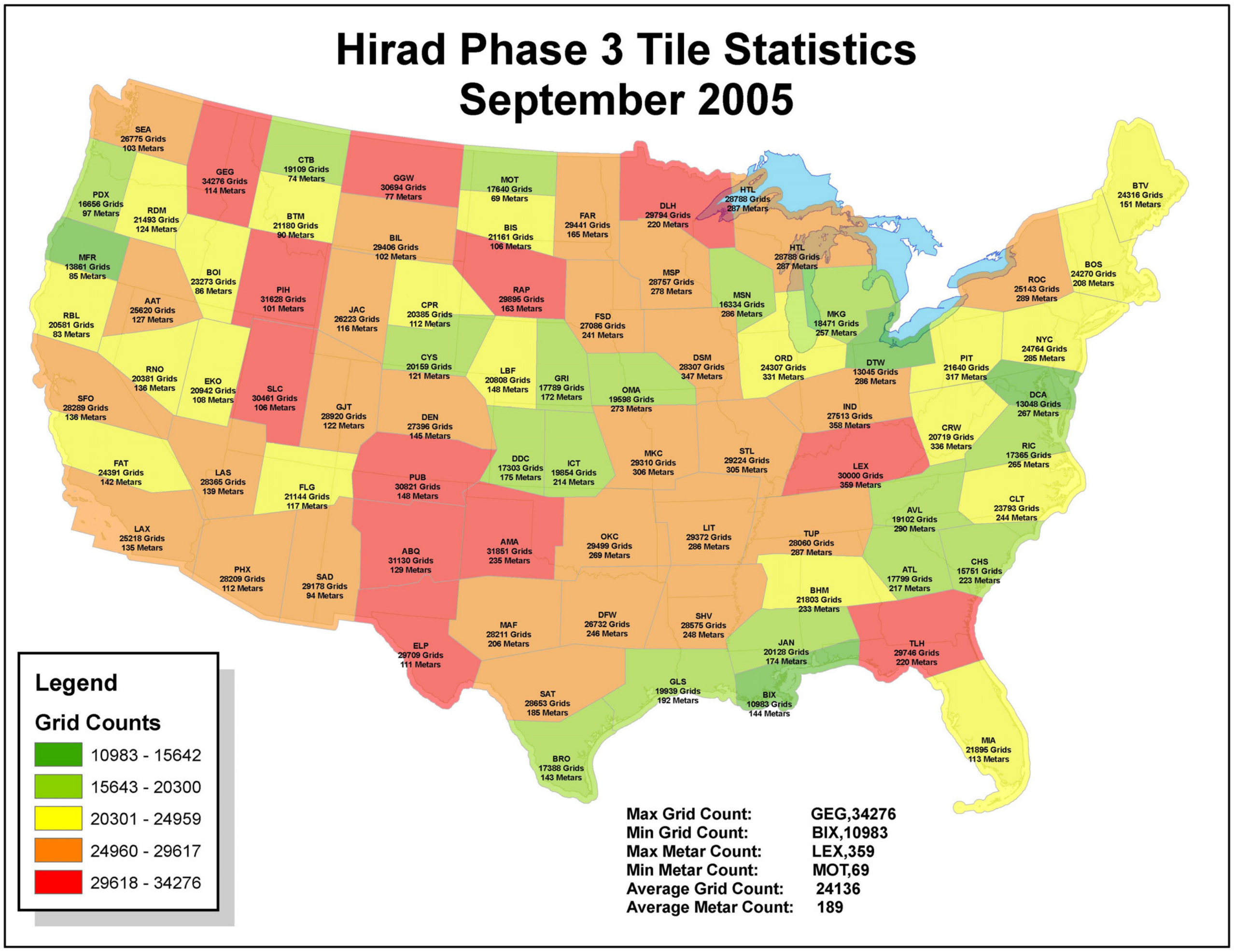

Fig. 6. The layout of DMA tiles used to produce the HiRAD phase 3 gridded output.

accumulation including accumulations of snow. Each HiRAD report will include running tallies of 1, 6, and 24-hour melted precipitation and snowfall. The snowfall integrations will use estimates of snow ratio following a method suggested by Cobb and Waldstreicher (2005). Early results from the 2005-2006 snowfall season are promising. In general, precipitation accumulations compared against gauge and WSR-88D precipitation accumulation estimates show HiRAD superior to WSR-88D estimations but lagging retrospective products such as the Stage IV hydrological analyses. This is not surprising considering the requirement of near real-time synthesis within HiRAD.

HiRAD gridded data will be available in March of 2006. The grid processing strategy will be divide-and-conquer on a large server farm. The CONUS 2.5 km grid is subdivided into 80 distinct tiles. The tiles are interlocking and adjacent (Fig. 6), but are irregularly shaped and of different sizes since the tile geometry is based on the distribution of demographic marketing areas or DMA's. A DMA subdivision is chosen since it is believed that tiles will be commonly used for visualization or applications of HiRAD data within individual markets; thus, it will be unnecessary in such local applications to incur the overhead of the entire gridded dataset. An example temperature plot of HiRAD phase III data based on the Denver, CO HiRAD tile is shown in Fig 7. This is compared with a simple contour of a METAR-only analysis in this same region.

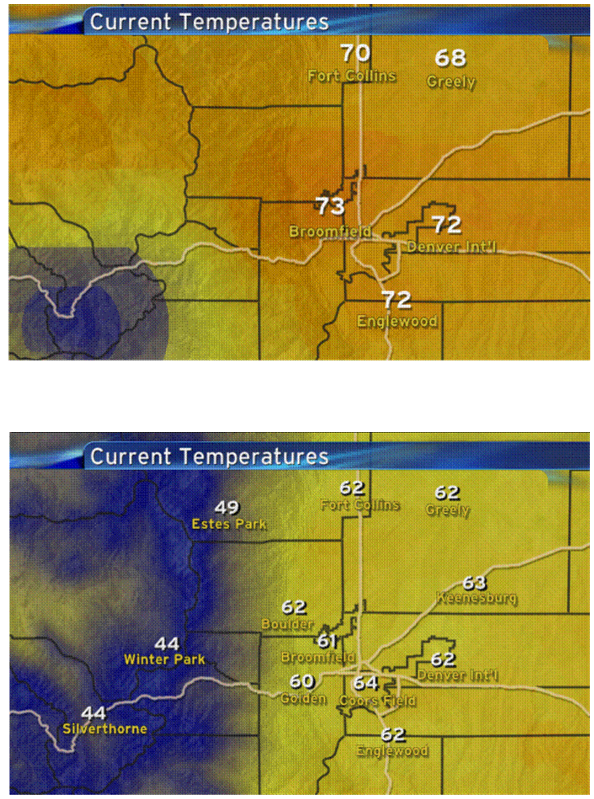

Fig 7. Examples of data plots and contours using conventional data sources (top) and the HiRAD phase III output grids (bottom).

6. Summary

This paper described the basic methodology, outcomes, and skill of a new system designed to provide consumer weather information of the current conditions at arbitrary locations. The system uses a variety of in-situ and remotely sensed observations including METAR reports, satellite data, radar observations, lightning strike data, high-resolution climatology and numerical weather prediction model output. In some regards, the system can be considered a data assimilation system. However, it has significant differences with traditional data assimilation systems designed to provide initial conditions for a numerical weather prediction system. The system has been deployed operationally in the United States for providing surface weather conditions for commercial use. Initial quality assessments have suggested the system produces results with statistical qualities approaching that of actual *in-situ* observational networks.

7. References.

Cobb, D. K. and J. S. Waldstreicher, 2005: A simple physically-based snowfall algorithm. Proceedings from the 21st Conference on Weather Analysis and Forecasting. Washington D.C..

Daly, C., R Neilson and D. Phillips, 1994: A Statistical-Topographic Model for Mapping Climatological Precipitation over

Mountainous Terrain. *J. Appl. Meteor.* Vol. 33, No. 2, p. 140-158.

Koval, J.S., J.G. Estupinan and J. Staudinger, 2006: Verification of a real-time system to estimate weather conditions at high resolution in the United States. Proceedings of the 22nd Conference on Interactive Information Processing Systems for Meteorology, Oceanography and Hydrology, Atlanta GA.

Oliver, M. A. and R. Webster, 1990: Kriging: a method of interpolation for geographical information systems. *Int. J. Geog. Info Sys.,* Vol. 4, No. 3, 313-332.

Rose, B.., N. McGillis, J.P. Koval, P. Neilley, andJ. G. Estupinan, 2006: Incorporating data-denial statistics into real-time quality control filtering of surface observations. Proceedings of the 22nd Conference on Interactive Information Processing Systems for Meteorology, Oceanography and Hydrology, Atlanta GA.