12.3 INCORPORATING DATA-DENIAL STATISTICS INTO REAL-TIME QUALITY CONTROL FILTERING OF SURFACE OBSERVATIONS

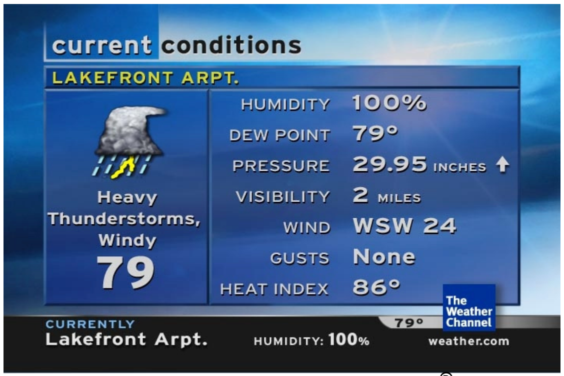
Bruce Rose[1], Neil McGillis[1], Joseph Koval[1], Peter Neilley[2], and Jeral G. Estupinan[1]

1. INTRODUCTION

The Weather Channel, based in Atlanta Georgia, has designed and operationalized a system called HiRAD (High Resolution Aggregate Data) for the purpose of producing real-time reports of current conditions within the Conterminous United States (CONUS). Within this domain, HiRAD estimates all common surface variables at high resolution in time and space and pays especially close attention to the present weather, visibility, precipitation accumulations of rain and snow, and cloud cover - since it is believed these data distinguish a HiRAD current condition weather report from other commonly available mesoscale observations that underreport these “sensible” weather data.

The HIRAD system is well described in Neilley and Rose (2006); additionally, HiRAD performance and skill validation are discussed in Koval et. al. (2006). HiRAD owes much of its conceptual framework to the US Weather Research Program’s Analysis of Record and Analysis of Moment projects (c.f. Horel and Colman, 2005). However, much of the details of this implementation of a real-time Analysis of Moment are quite novel and have not been previously attempted in an operational setting.

As described in Koval et. al. (2006), the key feedback of HiRAD accuracy or skill is determined using data-denial analysis. The HiRAD system has the ability to run in a data-denial mode. A shadow version of the system runs once per hour; in this shadow run, input METAR observations are systematically withheld from the core HiRAD analysis. In this way, a HiRAD output location such as KBOS or KATL become completely “synthetic” yet they are analyzed, derived, and output at identical times to the true production system (using all available METAR observations). It is this case-by-case comparison of the production versus data-denial versions of the system that produce validation statistics that, in turn, can be used to infer the total HiRAD error field; that is, the expected or estimated error at points where no corroborating surface observation or ground-truth exists.



current conditions

LAKEFRONT ARPT.

HUMIDITY 100%
DEW POINT 79°
PRESSURE 29.95 INCHES ↑
VISIBILITY 2 MILES
WIND WSW 24
GUSTS None
HEAT INDEX 86°

Heavy Thunderstorms, Windy
79

CURRENTLY Lakefront Arpt. 79° HUMIDITY: 100% weather.com

Figure 1.1 – *A “Local on the 8s®” current conditions page seen on the The Weather Channel cable network.*

Surface observations or current conditions (see figure 1.1) have always been a staple of The Weather Channel direct-to-consumer weather information products. For as long as these data have been mined, value-added and published – the quality control of the incoming observational data has been a thorny issue.

A crucial need for any assimilation system is robust quality control and filtering of incoming observation reports. While the incidence of bad observations is low with such data sources as ASOS and AWOS METAR reports, appreciable and egregious observation errors do occur across every important variable. These errors can be very large or quite persistent (i.e. the bad reports continue over many hours or even days); conversely, they can be subtle and very difficult to discriminate from so-called good observations. Fixed interval checks, spatial buddy checks, and temporal consistency checks are all well documented techniques for filtering bad observations (c.f. the description of the

Corresponding Author address: Bruce Rose, The Weather Channel, 300 Interstate N. Parkway, Atlanta, GA 30339. e-mail: brose@weather.com

[1] The Weather Channel, Atlanta, Georgia.
[2] Weather Services International Corporation, Andover, Massachusetts.

Forecast Systems Laboratory's MSAS/RSAS Surface Analysis System at http://www-sdd.fsl.noaa.gov/MSAS/masa_descrip.html). However, each approach has inherent weaknesses or involves costly calculation overheads – especially when measured against the desire of rapid processing and near real-time publication of the results.

The following sections describe a simple and computationally efficient quality control technique based on data-denial statistics accumulating in the HiRAD system. We present some results from nine months of operational use, and discuss the future directions and areas of continued work on quality control and observation filtering within the HiRAD environment.

2. OBSERVATION REPORTING ERRORS

The incidence of errors in METARs reported within CONUS is quite low; far less than 1% of total reports. For a typical example, on 3-NOV-2005 we ingested 75,824 total observations of which 139 were flagged or rejected in HiRAD for a filter ratio of 0.18%. The errors are more common with AWOS sensors compared to ASOS. The ASOS reporting errors are tracked and actively maintained by NWS's AOMC (ASOS Operations and Monitoring Center). A process is in place to open a "trouble ticket" on a suspicious or clearly erroneous sensor report for an ASOS. The turn-around time for resolution of reported problems is generally 1 day to 1 week.

ASOS sensor suites do self-diagnosis and add a special flag to METAR reports when the ASOS software detects a malfunction or senses it is in a degraded state. However, the diagnostic flag or reporting is most often non-specific providing only a simple on or off state. It means literally that something may be wrong with the myriad information typically reported in a METAR bulletin. More specific diagnostic information embedded within the METAR format would be beneficial (for example, as an indication of the specific sensor experiencing problems) and would provide the opportunity for additional downstream filtering or quality-control actions. Moreover, the ASOS sensor platform is capable of self-diagnosis but will often fail "open": meaning measurements that are clearly erroneous are allowed to continue transmitting until there is some human intervention to block the spurious reporting.

Within the AWOS network, no known standardized process for reporting problems on individual AWOS reports exist. The individual AWOS airfield or location must be contacted and – in many cases – we find that really nothing can or will be done to address a problem sensor. In the worst cases, specific problems can and do persist for weeks or months.

Military installations or bases often provide partial or total manual observation reports. In general, these reports are of high quality and reliability. The most prevalent type of error found at military installations appears to be simple typos. Most are completely isolated, but egregious. For some reason, wind speed coding errors seem to be most common within this group of stations.

In sum, the downstream quality control of observations is quite necessary since a small number of random but large-magnitude reporting errors occur on a daily basis.

3. METHODOLOGY

TWC recently completed a functional enhancement to HiRAD that includes a novel observation filtering technique called data-denial enhanced quality control or DDEQC. In this approach, interval checks for bad observations are distinct for each observation point. The variable intervals are based in part on the evolving database of data-denial results accumulating in the HiRAD system. These data-denial results can be aggregated as standard mean errors for a given point and time based on the accumulated knowledge of the nominal separation between the data-denial and production outputs of the observations. Thus, interval-based filtering for a point with large data-denial standard errors will result in conservative filtering or flagging of a bad observation. Conversely, a point with low data-denial standard error signals the system to be more aggressive in withholding a suspicious observation value.

Currently, this DDEQC approach is used for temperature, dew point, wind speed, and wind gusts and acts on all surface observations flowing into the HiRAD system. Results to date are encouraging. The DDEQC approach does not replace all other observation quality monitoring techniques, but can be a useful additional test when judging observation quality and confirming suspicious observations.

The incorporation of data-denial information into the quality control of HiRAD input observations is straightforward and can be summarized as follows:

1. Data denial statistics for temperature, dew point, wind speed, and wind gusts are saved for a period of one month for all regularly reporting METAR stations (about 1,400 in CONUS). There is currently no partitioning by time of day or other classification.

2. Each month a simple standard error is computed for each station and each variable (Ψ) according to equation 1.1-1.3, where σ is standard deviation, subscript *i* refers to an individual sample, superscript *DD-Prod* refers to the difference between the data-denial and production outputs, and the overbar refers to the sample mean (at least 72 cases are required to produce a standard error statistic and mean). This standard error is substituted into the definition of 95% confidence interval (assuming normal distribution) to determine the data-denial error for a given point (the positive root of Eq. 1.3). The data-denial error is clipped at some minimum value to prevent overly small quality-control thresholds from developing in the automated system.

$$\sigma_{DD}^{\Psi} = \sqrt{\frac{\sum_{i=1}^{n}\left(\Psi_i^{DD-\mathrm{Pr}od} - \overline{\Psi}^{DD-\mathrm{Pr}od}\right)^2}{n-1}} \qquad \text{Eq. 1.1}$$

$$StdError = \sigma_{DD}^{\Psi} n^{-1/2} \qquad \text{Eq. 1.2}$$

$$95\%CI = \overline{\Psi}^{DD-\mathrm{Pr}od} \pm \left(1.96 \times StdError\right).. \qquad \text{…Eq. 1.3}$$

3. As actual observations flow into HiRAD, real-time comparison is made between the RUC grid estimation and the observed value, *RUC-Obs*, according to Equation 1.4. This absolute difference is compared to the data-denial error for the point in question, times some scaling factor (ς_Ψ) which is determined from experimentation. The scaling factor lies between 6-10 for all variables.

$$\left|\Psi_{Actual}^{RUC-Obs}\right| < 95\%CI \times \zeta_{\Psi} \qquad \text{Eq. 1.4}$$

4. If the test in Equation 1.4 fails, the observed value is filtered or thrown out and is withheld from the downstream assimilation and analysis.

5. In the case of wind gusts, the filtering rules are more complex and the gust is constrained to be some multiple of the sustained wind, or is constrained to be related to the sustained wind – amongst other criteria.

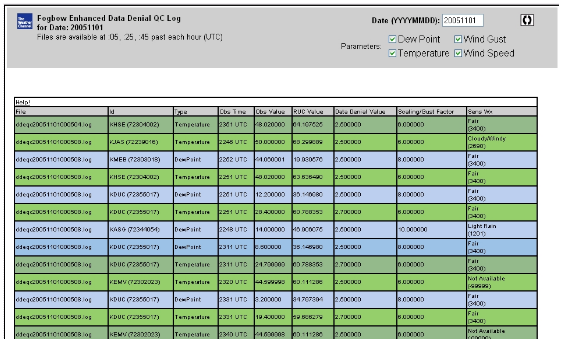
6. Another common exception is that the scaling factor is relaxed under certain weather conditions such as thunderstorms. This protects against erroneous withholding of temperature observations under conditions of strong local cooling brought about by convective precipitation.

The largest standard errors for temperature are seen in mountainous terrain and near coastlines. The month to month variation does not appear to be large for most points, so there is some sentiment that it may be unnecessary to refresh all of the standard error metrics on a monthly basis.

However, there is appreciable difference of standard errors for an individual point with respect to time of day. The standard error appears to peak in the early morning hours and finds a minimum value in the late afternoon hours. This diurnal variation is also seen with the wind speed standard errors. Variations in the standard error of dew point temperature appear to be insensitive to time of day for the period studied herein.

It is believed, but unproven, that the RUC skill in estimation of near-surface temperature and wind speed is best when the atmosphere is well-mixed. This gives low static stability so low and mid-level thermal and kinematic properties that are well predicted by the RUC numerical simulations are better communicated to the surface at times of strong daytime mixing. Given this behavior it may prove worthwhile to provide hourly standard errors from the data-denial statistics collection.

In this way, observations might be filtered differentially as the confidence in the RUC estimations vary through the day. This has yet to be attempted but would likely improve the hit ratio of filtering bad observations over good.



The Weather Channel

Fogbow Enhanced Data Denial QC Log
for Date: 20051101
Files are available at :05, :25, :45 past each hour (UTC)

Date (YYYYMMDD): 20051101

Parameters: ☑ Dew Point ☑ Wind Gust ☑ Temperature ☑ Wind Speed

Help!

File	Id	Type	Obs Time	Obs Value	RUC Value	Data Denial Value	Scaling/Gust Factor	Sens Wx
ddeqc20051101000504.log	KHSE (72304002)	Temperature	2351 UTC	48.020000	64.197525	2.500000	6.000000	Fair (3400)
ddeqc20051101000508.log	KJAS (72239016)	Temperature	2246 UTC	50.000000	68.299889	2.500000	6.000000	Cloudy/Windy (2690)
ddeqc20051101000508.log	KMEB (72303018)	DewPoint	2252 UTC	44.060001	19.930576	2.500000	8.000000	Fair (3400)
ddeqc20051101000508.log	KHSE (72304002)	Temperature	2251 UTC	48.020000	63.636490	2.500000	6.000000	Fair (3400)
ddeqc20051101000508.log	KDUC (72355017)	DewPoint	2251 UTC	12.200000	36.146980	2.500000	8.000000	Fair (3400)
ddeqc20051101000508.log	KDUC (72355017)	Temperature	2251 UTC	28.400000	60.788353	2.700000	6.000000	Fair (3400)
ddeqc20051101000508.log	KASG (72344054)	DewPoint	2248 UTC	14.000000	46.906075	2.500000	10.000000	Light Rain (1201)
ddeqc20051101000508.log	KDUC (72355017)	DewPoint	2311 UTC	8.600000	36.146980	2.500000	8.000000	Fair (3400)
ddeqc20051101000508.log	KDUC (72355017)	Temperature	2311 UTC	24.799999	60.788353	2.700000	6.000000	Fair (3400)
ddeqc20051101000508.log	KEMV (72302023)	Temperature	2320 UTC	44.599998	60.111286	2.500000	6.000000	Not Available (-99999)
ddeqc20051101000508.log	KDUC (72355017)	DewPoint	2331 UTC	3.200000	34.797394	2.500000	8.000000	Fair (3400)
ddeqc20051101000508.log	KDUC (72355017)	Temperature	2331 UTC	19.400000	59.686279	2.700000	6.000000	Fair (3400)
ddeqc20051101000508.log	KEMV (72302023)	Temperature	2340 UTC	44.599998	60.111286	2.500000	6.000000	Not Available

Figure 3.1 – *An internal web browser log of real-time observations filtered according to the DDEQC criteria.*

A daily log of flagged observations is produced by HiRAD (see figure 3.1) The most commonly flagged observation is dew point; the majority of the bad dew points are found on AWOS platforms. Temperature is less commonly flagged followed by wind speed or gust. The final disposition of a suspicious or flagged observation can only be determined, retrospectively, via inspection and research on a case-by-case basis. It is sometimes impossible to determine if an observation is in error given the noisy nature of near-surface quantities such as temperature and wind. This implies that the determination of gross tracking statistics, such as the hit/miss ratio for DDEQC filtering, are difficult and time consuming to derive.

4. DISCUSSION AND FUTURE WORK

The greatest weakness to date is the inadvertent filtering or withholding of good observations. The hit ratio for DDEQC is below 50% on most days. That is, more than 50% of the withheld observations are ultimately good observations. Relaxing the filter constraints via manipulation of the scaling factors for each variable can reduce this false-alarm number, but this then increases the risk of bad observations flowing into the HiRAD analysis system.

The end result of a "bad" filter of an incoming METAR observation is that the observation is withheld from the downstream analysis and estimation techniques. It rarely results in a missing HiRAD report. Instead, the report is synthetically produced from the remaining inputs – and is often indistinguishable from nearby "good" reports. It is this behavior that leads us to err on the side of withholding when an observation value is flagged as suspicious or problematic.

We believe the scaling factors are tuned appropriately, and the DDEQC is nearly optimized. This leaves two areas of additional work to be considered:

1. As previously discussed, resolve the standard error statistics for each variable and location to time of day. The data suggests a significant diurnal variability of the standard error fields for temperature and wind speed. The DDEQC technique could benefit from partitioning the error statistics by time of day - at least for these two variables.

2. Build additional real-time filters along the lines of MSAS/RSAS, but adapted for a real-time and fully automated system. We believe a spatial buddy-check as a filtering test might prove most valuable and is the highest priority in future development. That said, we are most sensitive to compute time needed for these additional quality control checks.

Finally, it is worth pointing out that HiRAD input surface observations are currently limited to ASOS and AWOS METAR's delivered via NWS FOS or NOAAPort services. We intended to aggressively pursue the addition of public or private *in situ* mesoscale network observations as inputs to HiRAD. While this remains a goal, we are proceeding more cautiously based, principally, on our experiences with the quality control and filtering of the [supposed gold-standard] ASOS and AWOS network data. Bad observations that remain undetected can seriously degrade the resulting analyses. The stakes are higher now that a single bad observation can impact large areas of resulting reports, which means our QC efforts must redouble compared to the previous status quo. Likewise, we must be cautious in introducing new input data sources to HiRAD.

For example, we prioritized the RAWS network of sensors as a potential input to HiRAD (c.f. DOI, Bureau of Land Management, 1997), since the data are freely available, many are located in alpine regions, and the stations are typically removed from population centers or regions of higher ASOS/AWOS density. One analysis that took place on incoming RAWS data was a comparison of RAWS temperature to ASOS temperature for a subset of RAWS locations virtually co-located with ASOS platforms (we found 29 of these).

The surprising result showed that RAWS temperatures diverged from their ASOS counterparts during sunny warm afternoons, but tracked closely to ASOS temperatures at night and under cloudy or rainy conditions. We can only conclude that the RAWS temperature sensors are poorly shielded or ventilated - at least for these sites. And we, thus, are reluctant to introduce this potential bias into the HiRAD analysis. We suspect other mesoscale networks will reveal separate but significant quality and accuracy issues when evaluated for inclusion as HiRAD inputs, and this assessment is on-going.

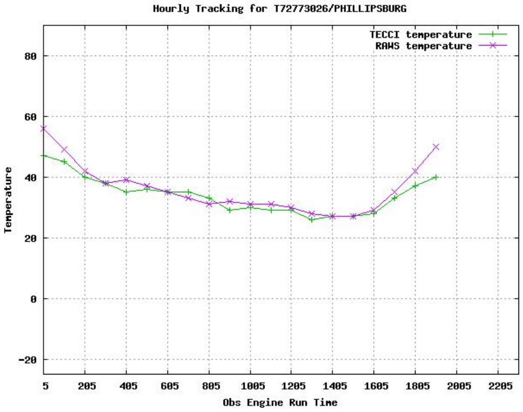


Figure 4.1 *– An individual comparison of a single day temperature trace for HiRAD (METAR-driven) versus RAWS report for Phillipsburg, Montana on 23-MAR- 2005. Note, the divergence of the temperature trace in the afternoon hours.*

5. REFERENCES

DOI, Bureau of Land Management, 1997: Remote Automated Weather Station (RAWS) and Remote Environmental Monitoring System (REMS) Standards. Boise, Idaho: RAWS/REMS Support Facility. 34 pp.

Horel, J. and B. Colman, 2005: Mesoscale Objective Analysis: An Analysis of Record?. Preprint. *9th Symposium on Integrated Observing and Assimilation Systems*. January, 2005. San Diego, California.

Koval, J., J. Estupinan, and J. Staudinger, 2005: Verification of a real-time system to estimate weather conditions at high resolution in the United States. Preprint. *22nd AMS Meeting on Interactive Information Processing Systems for Meteorology, Oceanography, and Hydrology*. January, 2006. Atlanta, GA.

Neilley, P. and B. L. Rose, 2005: A real-time system to estimate weather conditions at high resolution. Preprint. *22nd AMS Meeting on Interactive Information Processing Systems for Meteorology, Oceanography, and Hydrology*. January, 2006. Atlanta, GA.