6A.7

USE OF SURFACE MESONET DATA IN THE NCEP REGIONAL GSI SYSTEM

Seung-Jae Lee*, David F. Parrish, Wan-Shu Wu,
Manuel Pondeca, Dennis Keyser, and Geoffery J. DiMego

NCEP/Environmental Meteorological Center, Camp Springs, Maryland

1. INTRODUCTION

The NCEP analysis system has gone through many changes since optimum interpolation (OI) became its main component in 1978 (Bergman 1979; Dey and Morone 1985; DiMego 1988; Kanamitsu 1989; Derber et al. 1991; Parrish and Derber 1992). Recently, the 3DVAR-Gridpoint Statistical-Interpolation (GSI) System was developed as the next generation global analysis system (Wu et al. 2002) which uses recursive filters in grid point space to model the action of the background error covariance matrix upon the spatial distribution of observation increments (Purser et al. 2003). Furthermore, the current unified global/regional analysis system was developed after realizing that it was fairly straightforward to modify the GSI for both global and regional applications.

The NCEP GSI analysis system can assimilate diverse kinds of observation data, such as synoptic, satellite, and radar data. Especially as the importance of and demand for real-time mesoscale analysis grows in the world of weather forecasting, near-surface data assimilation becomes one of the challenges to conquer. This mesoscale surface analysis is expected to greatly contribute to improvements in short-range model forecasts and forecasters' analysis of mesoscale weather. Fortunately, the United States has high-resolution observation networks over land, and surface observation data are abundant. For example, the U. S. meso-network systems measure and provide useful information on the environment at the size and duration of mesoscale weather events.

In this study, we carry out assimilation experiments of surface mesonet data in the NCEP regional GSI system. The effort is focused on understanding the characteristics of innovations (observed–guess) of the surface mesonet data. Single time analyses are conducted, and modifications to background errors are considered.

* *Corresponding author address*: Seung-Jae Lee, Numerical Weather Prediction Division, Korea Meteorological Administration, 60-18 Shindaebang-dong, Dongjak-gu, Seoul, 156-720, Republic of Korea; e-mail: sjlee@kma.go.kr

2. DATA

2.1 *The Surface Mesonet Data*

The PREPared BUFR (PREPBUFR) file is a special application designed at NCEP to provide user-friendly access to the WMO-BUFR files through a series of FORTRAN subroutines in a machine independent BUFR library. In the current operational regional analysis and forecast system's PREPBUFR files, all mesonet data are included but the observation error for all mass (temperature, surface pressure, moisture) and wind observations is set to missing. The mass data for a mesonet report is stored under PREPBUFR report type 188 and the wind data is stored under PREPBUFR report type 288.

Because the observation error is missing, an extra "layer" of quality control is added in the PREPBUFR processing. This quality control sets the quality marker to "9" for surface pressure, temperature, specific humidity and wind. Previously the quality markers for these data were either "1" (for good), "2" (for neutral or not checked) or "3" (for suspect). A higher quality marker indicates a lower observation quality. All quality markers of 4 and higher (4-15) are considered "bad" (for various reasons) and the observation will not be assimilated.

We at NCEP currently do not perform any automated platform-specific quality control on mesonet (or any surface) data. There are good, neutral, suspect, or bad quality markers on the data as it comes in based on the FSL-MADIS quality control. In addition, NCEP could (but currently doesn't) place manual quality markers on the mesonet data (either "0" for keep or "14" for purge). This would be done by either putting reports on a reject list which might be updated monthly or on a report-by-report basis by the NCEP Senior Duty Meteorologist (SDM) who works in NCEP Central Operations. Also, we perform some gross checks and flag data that are outside reasonable limits, data with missing latitude or longitude, etc. Since the observation error is missing in operations, all mesonet data get a quality marker of "9" and the analysis then skips over them even though they are in the operational PREPBUFR files.

In this study, we modify the observation error file to assign the mesonet observations in PREPBUFR report types 188 and 288 the same observation error as for METAR mass and wind observations in report types 187 and 287, respectively. Clearly, this procedure is somewhat arbitrary, but it is adopted due to the lack of information regarding the observation error for mesonet data. NOAA FSL, for example, assigns to mesonet data the error values used for other surface data but inflates it by a factor of 1.5. In our special runs, we are making PREPBUFR files identical to the operational ones, except the observation errors for the mesonet data are now not missing. In this case, the extra layer of quality control does not occur. The original quality markers are retained for the mesonet data, so the analysis will assimilate all mesonet observations that do not have a bad quality marker due to the original MADIS quality control. The time window over which data collection is performed is +/- 1.5 hours, and the 3DVAR analysis experiment is valid for t=0h.

2.2 *Modification to Background Error Statistics*

Until now, the background error for regional assimilation has been a downscaled version from the background error derived from the global model. In the existing approach, horizontal scales are estimated from derivatives, but in the new approach, these are estimated using auto-covariances (W.-S. Wu 2005, personal communication). Both approaches use the NMC method, with global model forecasts for the existing approach and regional model forecasts for the new approach. For the global derived background error, the correlation length scales and background variances are a function of latitude and vertical. In the new background error, the scales of the structure function vary in the vertical but are no longer latitude dependent for statistical robustness.

It may be inappropriate to use global background error for the assimilation of high-resolution surface data. Therefore, in this study, the effect of regional versus global background error was examined in addition to sensitivity of regional GSI system to mesonet data. Table 1 shows the experimental design based on what has been described above. WRF 8 km analyses are used as background fields. There are three domains available for the WRF-NMM 8 km model: Initial fields for the western, central, eastern domains are at 06, 12 and 18 UTC respectively. Experiments are carried out for each model domain.

3. RESULTS AND DISCUSSION

3.1 *Analysis Comparisons*

Figure 1a shows the result for a single analysis on the western domain, valid 0600 UTC 14 Feb 2005. According to satellite and radar images (Fig. 2a), the weather was clear over most of the western USA domain at this nighttime. There are large differences between the analysis and guess fields in this case. These results are very much what one would expect, and stress the problems we currently have with attempting to use surface temperature data. For this case, because of the clear nighttime inversion over much of the domain, the observations are consistently colder than the model, leading to the large negative analysis – guess at the surface. Part of the reason this is undesirable for the current data assimilation is that temperature increments are coupled with wind increments in approximate geostrophic balance, and large wind increments are created at middle levels in the troposphere by large surface temperature increments. This is because the vertical correlation length for the balanced part of the analysis increment is constant (or slowly varying only in latitude) and is rather large. Another factor is that the horizontal correlation length scale is rather large, leading to a smooth large scale temperature increment.

Figure 1b shows the result for single analysis on the central domain, valid 1200 UTC 10 Mar 2005. It can be seen that the analysis field has smaller and detailed structures as mesonet data are added. Noticeably, there are positive analysis increments in the northern-central region where the low pressure system is located (Fig. 2b). When the regional background error statistics are used, the overall pattern is similar to the control run but the amplitude is somewhat intensified. It is believed to be the results of smaller vertical structures in the regional background error covariances.

The mean temperature profiles computed from the guess and analysis fields, and also some local soundings confirm the occurrence of low level nighttime inversions in the case of Figs. 1a and b: The inversion depth was about the 6^{th} or 7^{th} lowest-sigma levels, which means the increments actually extended to approximately 920 to 900 hPa levels.

Figure 2c shows the daytime (1400 LST) eastern case where a low-pressure system is located over the eastern coast. Although we cannot know much about the background until we do some assimilation experiments, even in the east coast case, the analysis - background is large (Fig. 1c). It just has smaller scales, reflecting the fact that there is a low with a large amount of local variation, whereas the west situation was characterized by clear nighttime stable surface layer, which at least in this case was very large in scale.

3.2 *Accumulated Statistics of Observation Innovations*

In order to identify the source of the large analysis increments, an attempt was made to look at the temperature increments (observed - guess) for the mesonet data. In the GSI, if the pressure of the observed surface temperature is less than the guess pressure at sigma level 1, then the guess temperature is interpolated in the vertical to the observed pressure. If the pressure of the observed surface temperature is greater than the guess pressure at sigma level 1, then the guess temperature is the value at sigma level 1. To see if the large apparent difference bias between observed and guess results from extrapolation outside the model domain, the observations are divided into those where (1) the observed pressure is less than the sigma 1 guess pressure, in which case the guess temperature is interpolated to the observed pressure; and (2) all remaining observations where observed pressure is larger than the sigma 1 guess pressure. We looked at the average of (observed - guess) for both cases.

Figure 3 displays the scatter diagram of observed versus first-guess surface temperature (˚C) during May 2005 for the night-time (0600 UTC) western, dawn-time (1200 UTC) central, and day-time (1800 UTC) eastern domains. Surface mesonet temperature data have a considerable amount of outliers compared with other land surface temperature data. The outliers imply not only bad observations but also local effects (Fig. 3a). Some stations can be seen to produce the same values regardless of model forecasts. This can mean quality markers placed on the data by the FSL-MADIS quality control are of little value not only for wind but also temperature data.

The slope and correlation coefficients (r^2) of land surface temperature data are good and similar in all three domains. However, in the case of synoptic sea surface data, they show a peculiar pattern in the western domain (Fig. 3d): a steep slope and very low correlation coefficients below 0.5.

As for the land surface temperature data, the nighttime western domain shows the worst RMSE among the three domains. In the case of synoptic sea surface data, the eastern domain shows the worst RMSE and the western domain the lowest correlation coefficient.

Figure 4 shows the mean Innovations for the surface temperature data as a function of surface pressure difference between model and observation for the month of May 2005. The nighttime western and central domains indicate a model warm bias. The western domain, in particular, shows a model warm bias of about +2.2 C at nighttime when compared with the eastern domain which did not show any obvious bias. The o-g statistics as a function of surface pressure difference in the eastern domain seems to indicate that the mesonet observations have small temperature bias (about +0.5 C). The number of stations as a function of surface pressure difference in the western domain revealed an asymmetric distribution that implies that there are many stations where the model surface is higher than the observation surface (not shown here). This could partly account for the model warm bias in the western domain at nighttime. In the other two domains, the distributions are comparatively axisymmetric.

Unlike the other two types of surface stations, surface mesonet stations are very dense, particularly around large cities (Figs. 5a, c, and e). Stations with large innovations are distributed uniformly in the nighttime western and central domains, while are mainly located in the large cities in the daytime eastern domain (Figs. 5b, d, and f). In the case of the eastern domain, it is 14 LST and the synoptic situation is characterized by many local and unstable situations with small-scale variation in daytime. In the case of the western (central) domain, it is midnight (dawn) and frequent large-scale inversion situation prevail. Therefore, it seems that the observations are not bad in western and central domains, but the model was perhaps in error in all domains as deduced from the homogeneous distributions of large (o-g) stations. These differences could also be the result of urban heat island effects (not contemplated in the model) or erroneous station groups.

4. SUMMARY AND FUTURE WORKS

In this study, we have assimilated surface mesonet data in the NCEP regional Grid-point Statistical Interpolation (GSI) using the same observation error as that adopted for METAR data within the WRF-NMM 3DVAR system. In the single-time analysis experiments, the analysis field was shown to contain smaller and detailed structures as mesonet data are added. The use of new regional background error statistics, not interpolated from global model error statistics, was shown to produce single analyses that differed more with the background.

To understand the characteristics of the observed - guess statistics for the surface temperature data, accumulated statistics of observation innovation have been evaluated for May 2005. According to the statistics, Surface mesonet temperature data were found to have a considerable amount of outliers compared with other land surface temperature data. The nighttime western and central domains indicated a model warm bias. The western domain, in particular, showed a model warm bias of about +2.2 C at nighttime when compared with the eastern domain which did not show any obvious bias. Stations with large innovations are distributed uniformly in the nighttime western and central domains, while they are mainly located in the large cities in the daytime eastern domain.

Future tasks include short assimilation experiments using mesonet data, the application of non-linear quality control to the mesonet data, the use of anisotropic background error covariances, and devising strategies to reduce the bias in the observation innovations.

5. REFERENCES

Burgman, K., 1979: Multivariate analysis of temperature and winds using optimum interpolation. Mon. Wea. Rev., 107, 1423-1444.

Derber, J. C., D. F. Parrish, and S. J. Lord, 1991: The new global operational analysis system at the National Meteorological Center. Wea. Forecasting, 6, 538-547.

Dey, C. H., and L. L. Morone, 1985: Evolution of the National Meteorological Center global data assimilation system: January 1982-December 1983. Mon. Wea. Rev., 113, 304-318.

DiMego, G. J., 1988: The National Meteorological Center Regional Analysis System. Mon. Wea. Rev., 116, 977-1000.

Kanamitsu, M., 1989: Description of the NMC global data assimilation and forecast system. Wea. Forecasting, 4, 335-342.

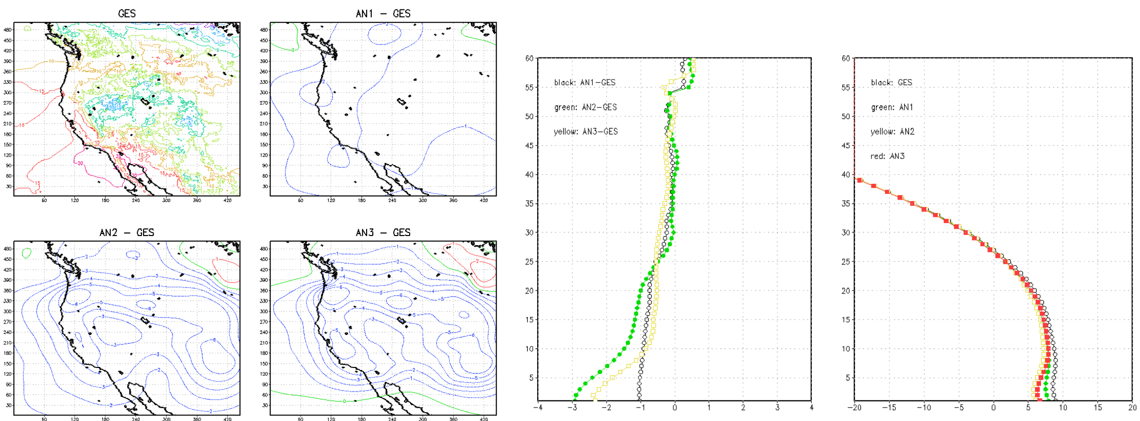
Parrish, D. F., and J. C. Derber, 1992: The National Meteorological Center's Spectral Statistical-Interpolation analysis system. Mon. Wea. Rev., 120, 1747-1763.

Purser, R. J., W.-S. Wu, D. F. Parrish, and N. M. Roberts, 2003: Numerical aspects of the application of recursive filters to variational statistical analysis. Part II: Spatially inhomogeneous and anisotropic general covariances. Mon. Wea. Rev., 131, 1536-1548.

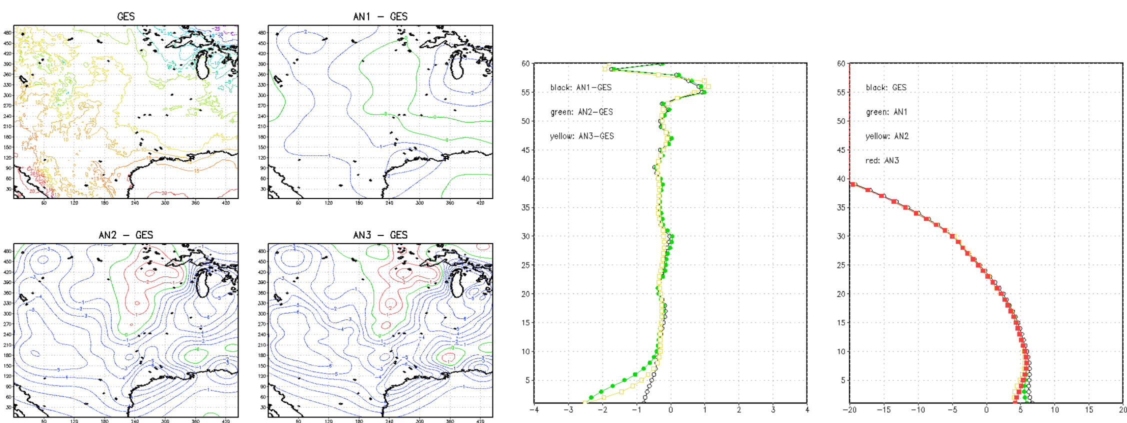
Wu, W.-S., R. J. Purser, and D. F. Parrish, 2002: Three-dimensional variational analysis with spatially inhomogeneous covariances. Mon. Wea. Rev., 130, 2905-2916.

Table 1. Experimental design for the analysis time 06 UTC 04 Feb, 12 UTC 10 Mar, 18 UTC 23 Mar 2005.

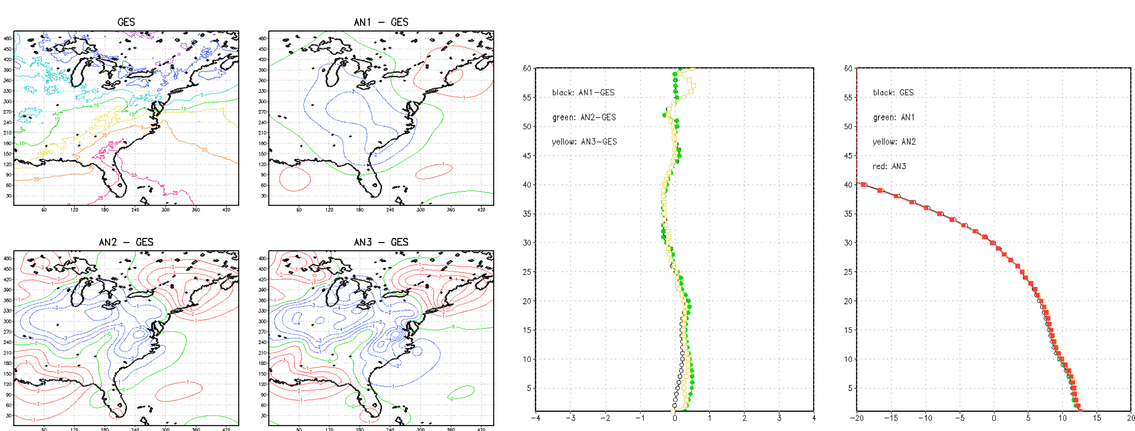
analysis	use of mesonet data	background error	Note
AN1	No	global	operation (control run)
AN2	Yes	global	impact of mesonet data
AN3	Yes	WRF-NMM	impact of new regional background error statistics



(a)

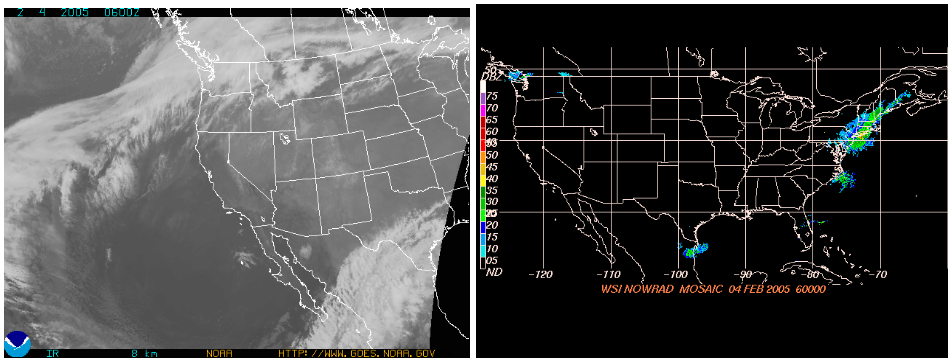


(b)

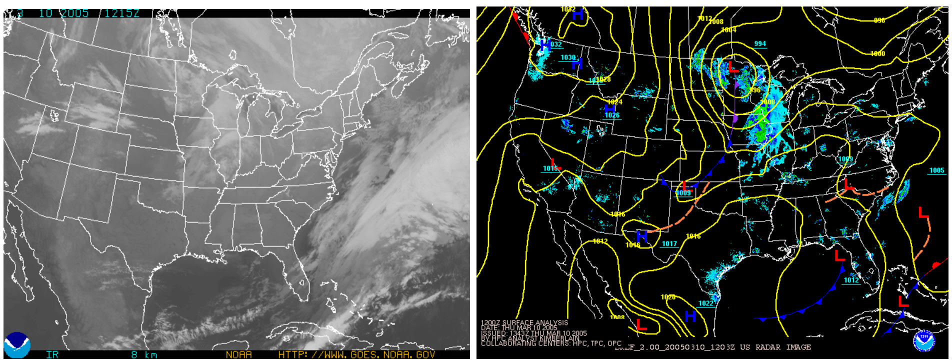


(c)

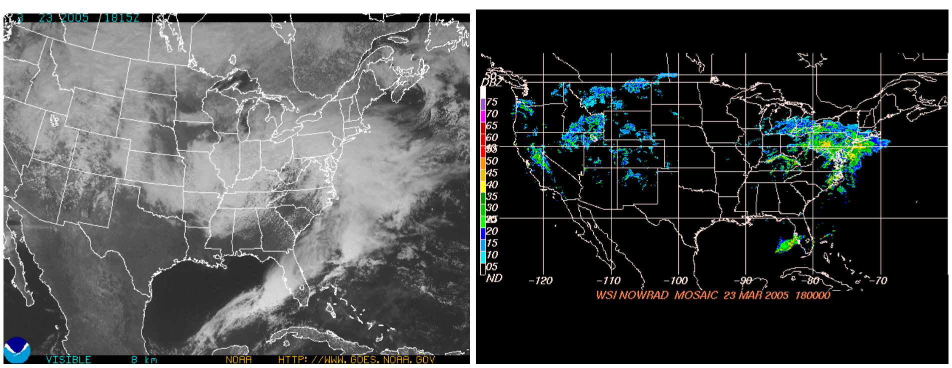
Figure 1. Horizontal fields (left four panels) and mean profiles (middle one panel) of analysis increments for temperature at (a) 0600 UTC 04 Feb, (b) 1200 UTC 10 Mar, and (c) 1800 UTC 23 Mar 2005. Right one panel is mean profiles of full guess and analyses.



(a)



(b)



(c)

Figure 2. Satellite and radar images (a) 0600 UTC 04 Feb, (b) 1200 UTC 10 Mar, and (c) 18 UTC 23 Mar 2005.

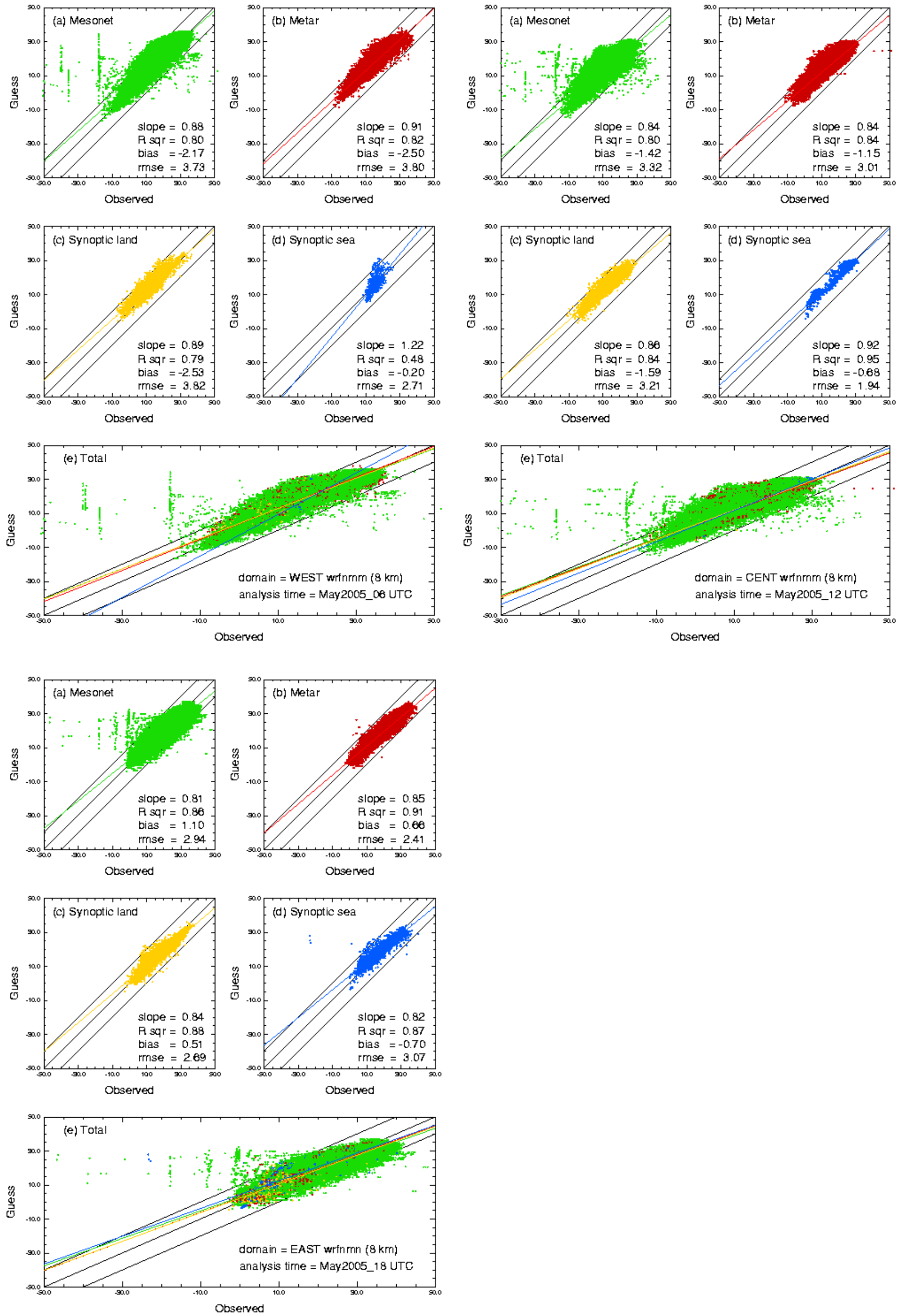


Figure 3. Scatter diagram of observed versus first-guess surface temperature (˚C) during May 2005 for the nighttime (0600 UTC) western (top left), early morning (1200 UTC) central (top right), and daytime (1800 UTC) eastern domains (bottom).

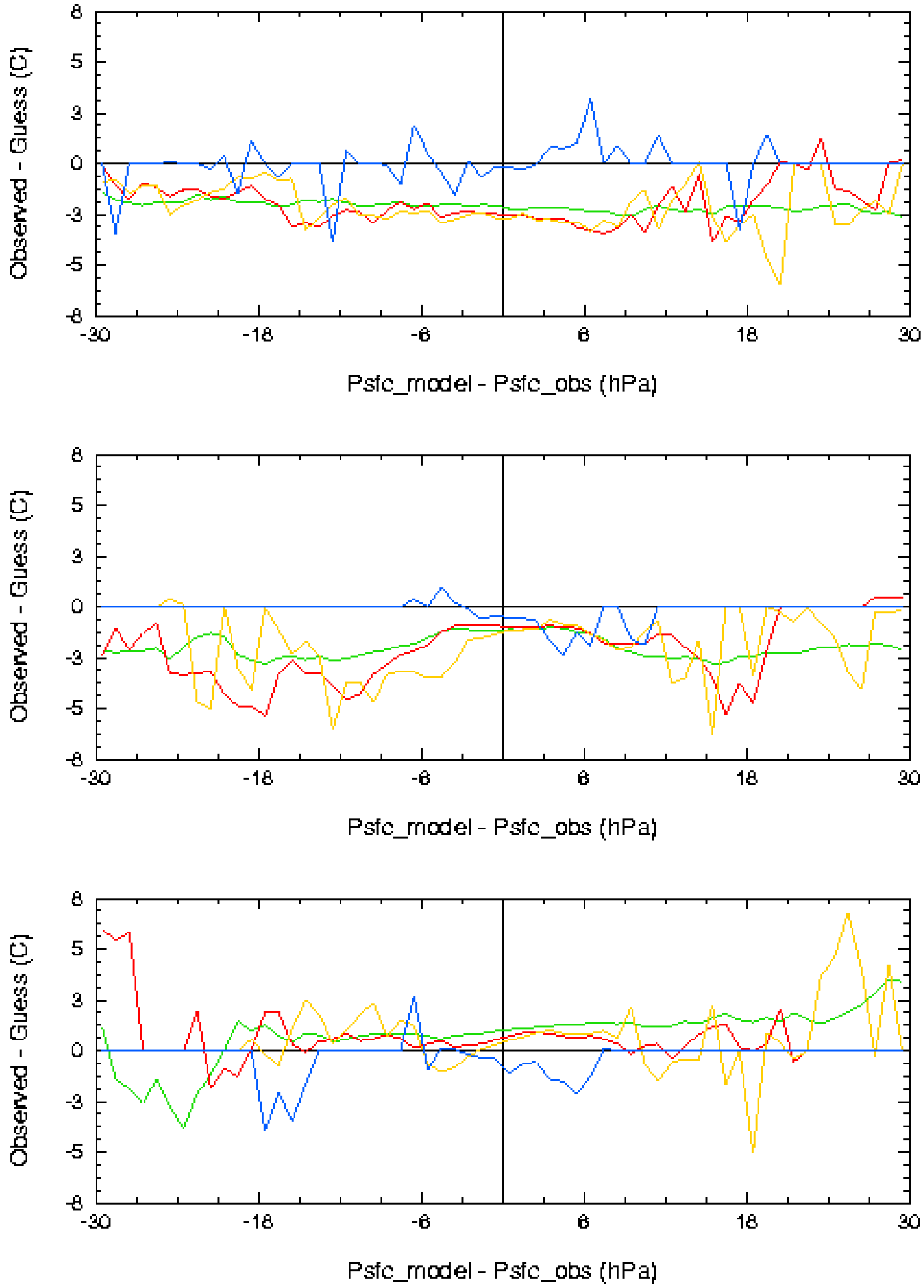


Figure 4. Mean Innovations of the surface temperature data as a function of surface pressure difference between model and observation during May 2005 for the night-time (0600 UTC) western (top), dawn-time (1200 UTC) central (middle), and day-time (1800 UTC) eastern domains (bottom). The color legend and time period are the same as Fig. 3.

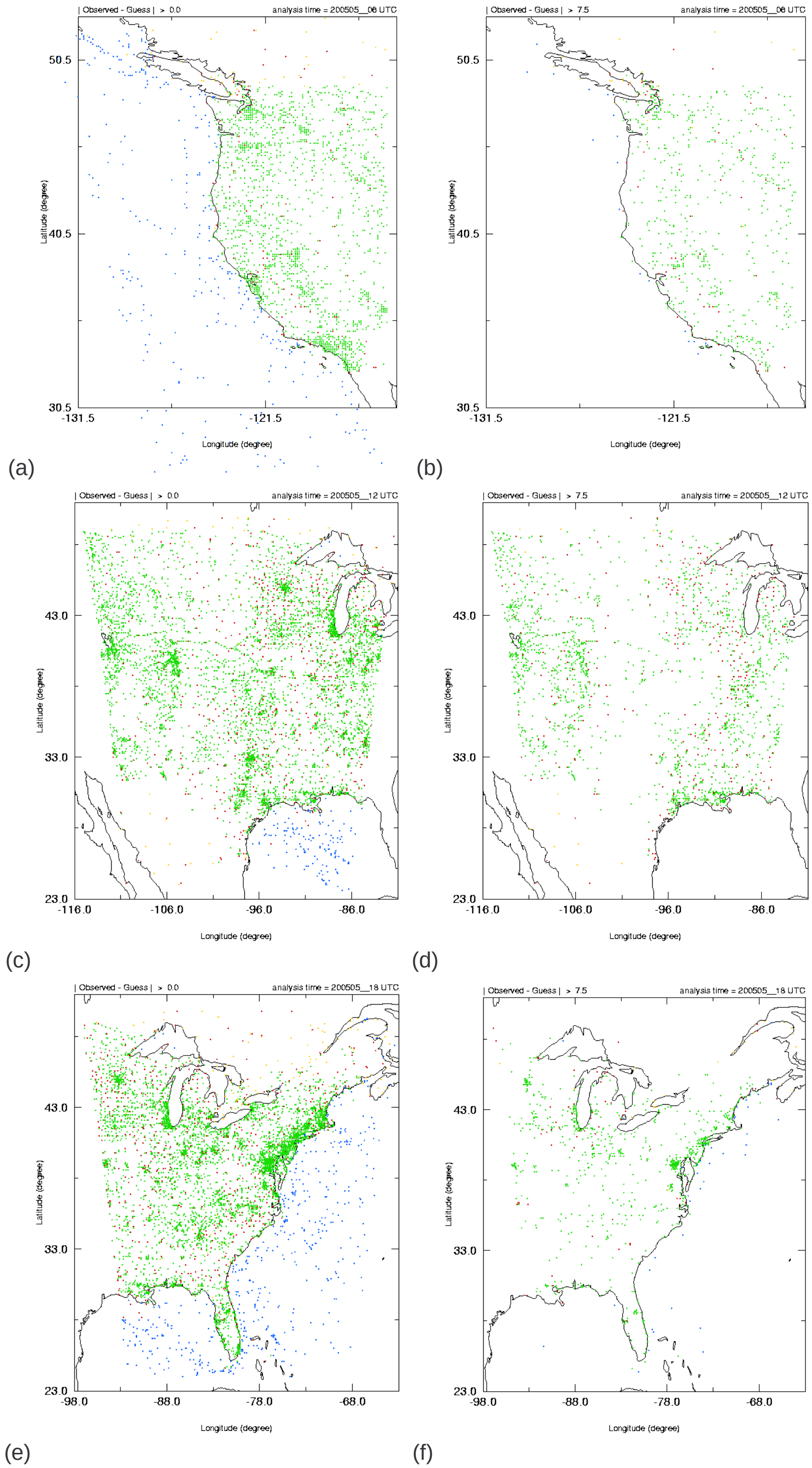


Figure 5. Stations which had large innovations in each domain during May 2005. The color legend and time period are the same as Fig. 3.