

J1.7 OBSERVATION-QUALITY ESTIMATION AND ITS APPLICATION IN THE NCAR/ATEC REAL-TIME FDDA AND FORECAST (RTFDDA) SYSTEM

Yubao Liu*, Francois Vandenberghe, Simon Low-Nam, Tom Warner and Scott Swerdlin
National Center for Atmospheric Research/RAP, Boulder, Colorado

1. INTRODUCTION

In the last three years, the National Center for Atmospheric Research (NCAR) and the Army Test and Evaluation Command (ATEC) have been developing a multi-scale (with grid sizes of 0.5 - 45 km), rapidly cycling (at time intervals of 1 - 12 hours), real-time four-dimensional data assimilation and forecasting (RTFDDA) system. By August, 2003, RTFDDA systems were customized and deployed to five Army test ranges, and to seven other regions to support specific missions of three other US government agencies. A Newtonian-relaxation-based "station-nudging" approach, by which all observations that are available in real-time are incorporated into a continuously running MM5 model, is employed to accomplish four-dimensional data assimilation. The nudging-based data assimilation weights each observation uniquely according to the observation time and location, and thus allows ingest of conventional and unconventional observations that are available at regular and irregular times intervals. The data sources incorporated include the traditional hourly surface (METAR, ship, buoy and special) reports and twice-daily upper-air rawinsondes. Also used are high-frequency measurements from various mesonets and special field experiments; wind profiler data from NOAA/FSL NPN profilers and CAP-Cooperative Agency Profilers; NOAA/NESDIS hourly GOES winds derived from IR, visible and water-vapor images; aircraft reports (ACARS/AMDAR) processed and disseminated by NOAA/FSL; and data from other non-conventional sources.

The "station-nudging" approach appears to alleviate some of the problems in mesoscale data assimilation and prediction. Another remaining problem is that data from different platforms have different instrument, sampling and processing errors and these errors may vary in time and space depending on the according to weather regime, instrument sitting, and geographic effect. In addition, when observations are analyzed onto a data-assimilation-model grid, representativeness errors appear, which can significantly affect the accuracy of analyses and forecasts. How to account for the overall effect of these errors (referred as to total observation error) is a critical problem in data assimilation. This is especially challenging because On the other hand, irregular time and space distributions of observations from the

non-conventional measuring platforms can make traditional data quality control approaches, such as buddy-checks and dynamic consistency check (as reviewed in next section), very complicated and computationally so expensive that it will hinder real-time usage. This will become particularly true because data volumes from these platform are increasing rapidly.

In this paper, a simple and efficient data quality-control (QC) procedure is described. The new QC module can grossly estimate the error of an observation, and thus is able to assign a quality level to individual observations. It is demonstrated that the RTFDDA data assimilation and forecasts are improved by using the data-quality information, obtained from the QC, procedure to weight each observation uniquely with confidence levels during the assimilation.

2. BACKGROUND REVIEW

Much progress in numerical weather prediction (NWP) on all scales have been achieved in the last three decades. Modern operational and research NWP models run with high resolution, complete dynamics and sophisticated physics parameterizations. Advanced filter schemes are employed to produce model initial conditions, using abundant data. Short-term (0 - 12 hour) forecasts are often used as the background for data analyses. The advances in the modeling technology and the computing capacity, for rapid forecast-analysis-cycling, significantly improve the model-background accuracy and hence the analyses based on it. Meanwhile, accurate short-term forecasts can be used to monitor the observation quality. Hollingsworth et al. (1986) point out that the 6 - 12 hour forecasts used as background for the operational analyses at ECMWF have an accuracy comparable to that of the radiosonde observation, for a study period of March and April 1984. With case studies, they showed that the ECMWF data analyses and 6 - 12 hour forecasts can be used to define bad radiosonde observations. Today's NWP forecasts are even more accurate. For example, 0 - 3 hour nowcasts are available in realtime from the rapidly cycling NCAR/ATEC RTFDDA system. Such nowcasts provide a superior basis for monitoring and quantifying observation errors.

Before introducing the new QC scheme, we briefly review observation error sources and the heuristic data QC approaches used in NWP. Essentially, two kinds of observation errors exist: instrument errors and representativeness (or

* Corresponding author address: Yubao Liu, National Center for Atmospheric Research, P. O. Box 3000, Boulder, CO 80307-3000. Phone: (303) 497-8211
Email: yliu@ucar.edu

sampling) errors. The instrument error is caused by sensor damage/calibration, uncertainties in variable-retrieving schemes, and human/electronic processing. These errors can either be systematic (bias error) or random (random error) or fatal error (nonsense observations). The representativeness error reflects the degree to which an observation represents the volume-average value. Representativeness errors are normally unbiased, and their magnitudes may depend on the volume size, the sensor's time average, the volume size which one wants the observation to represent, and the weather state. All measurements are subject to these errors, and we lack enough information to accurate each of these errors. Although most of these error sources seem to be platform-dependent, as assumed by most NWP centers, observation errors from the same platform type may vary with weather regimes and instrument siting.

At most research and operational NWP centers, two steps are involved in data analysis/assimilation: observations are firstly quality-controlled, and then they are used in statistical analysis. At the data QC stage, observations which do not satisfy the specified criteria in one or more checks are considered as "bad" observations and removed. Then, at the analysis stage, statistics of platform-dependent observation errors and models are used to define the relative weight of observations and background. Most NWP centers employ the following approaches to define "bad" observations: 1) validity and static consistency checks to define "impossible observations"; 2) "error-tolerance" checks by which observations are compared to 6 - 12 hour model forecasts, 3) "buddy-checks" where "neighboring" observations are compared with each other; 4) time-consistency check by which temporal abnormalities are found; and 5) dynamic/physical consistency checks where observations are examined for their dynamic/physical consistency with their environment,

Although the traditional data QC procedure, and the statistical-error-based data filters, work satisfactorily in NWP practice, it is obvious that the "sharp-line" drawn between "bad" and "good" in the QC schemes is arbitrary. For example, given two temperature observations: one is 0.1 C above and the other is 0.1 C below the "error-max" cut-off line. It is very arbitrary to drop the one that is 0.1 C over the error-max and to keep the other one that is 0.1 C under the error-max. It is quite possible that a "bad" observation that fails the QC checks and is dropped out, is a "good" observation. Furthermore, typical QC schemes can not properly deal with the observation errors that vary in response to weather regimes and geographical effects, even for the observations made from the same measuring platforms. This is particularly true when applying these schemes in mesoscale modeling systems, such as the NCAR/

ATEC RTFDFA system, since on the mesoscales observations are relatively sparse and weather systems can change greatly in time and space. Also, the diverse non-conventional observations used may have very different errors. Therefore, it is necessary to refine the QC schemes in order to better quantify and use the observation errors.

Statistical and variational analysis, 3DPAS and 3DVAR, are the most common analysis schemes employed for data assimilation and initialization of models in the modern operational NWP center. These schemes heavily rely on the errors of observations and model. The observation and model errors are normally estimated based on historical observations and model forecasts. In these schemes, the historical statistical background errors are assumed to be the error of the individual background during each analysis, and similarly, the historical statistical observation error is used to approximate the error of individual observations. These assumptions are faulty, especially when applied to mesoscale modeling. In fact, the concept "error of the day" relative to backgrounds for data assimilation was raised more than a decade ago. Basically, the "error of the day" concept says that the true (or the best estimate of) error of the current background should be used, and that the statistical errors obtained based on historical model forecasts are not sufficient. Similarly, we point out that an accurate data assimilation system also needs the "error of the day" for observations - the error of each individual observation. Like the background errors, observation errors for a given set of observations will always differ from the historical statistics error. In addition, since the representativeness error of observations is a function of the size of the analysis grid box and the weather, the representativeness error is an important part of the "error of the day" of observations, which can not be properly addressed by the traditional data QC and analysis schemes.

Defining the "error of the day" for either the background or the observations is a very challenging problem. Although there are many research activities related to using the ensemble prediction approach to describe the "error of the day" for backgrounds, such techniques have not been used in operations. The "sharp cut-off" method used to remove "bad" data in the traditional data QC schemes appears to inconsistent with the concept of the "error of the day" for observations.

In the following sections we describe a new data QC scheme developed in conjunction with the Newtonian-relaxation-based rapidly cycling operational NCAR/ATEC RTFDFA system. The scheme is constructed to examine and estimate the relative quality of each individual observation. We will also demonstrate how to use this "error of the day" information in the station-nudging-based data assimilation,

and evaluate its impact on the RTFDFA analyses and forecasts using case studies.

3. CONCEPT AND DESIGN OF A NEW QC SCHEME

An NWP system that “cold-starts” from three-dimensional analyses has forecasts that experience initial dynamic and physical adjustments due to the inconsistencies between the analyses and the model. This is commonly referred to as “spin-up problem”. In contrast, a FDDA-based analysis and forecast system suffers from the “spin-up problem” because the 4-D process fits the model to observations, and maintains the model balance at the same time. In an FDDA and forecast system, the shorter the forecast, the more accurate the forecast. Analyses are most accurate, 1-h forecast is better than the 2-h, the 2-h is better than the 3-h and so on. The NCAR/ATEC RTFDFA system is one such system that produces, in real-time, 4-D dynamically and physically consistent analyses and forecasts. To illustrate this property, Fig. 1 presents an example of the verification of the RTFDFA system running at the Army’s Aberdeen Proving Ground during August, 2003. It can be seen that the analysis and short-term forecast errors are comparatively small, and the forecast accuracy gradually decreases with forecast length. The RTFDFA 0 - 3 hour forecasts provide a fairly accurate background that can be used to examine the overall error (referred to as total error, hereafter) of real-time observations and model backgrounds, in the sense of “error of the day”. (note: 0h forecasts are FDDA analyses). Furthermore, by comparing the observation-background differences with the statistics of the model and observation errors it might be possible to scale the quality of each observation. The new RTFDFA QC procedure was designed according to this concept.

Firstly, let us assume that the RTFDFA analysis and 1 - 3 hour forecasts are perfect (i.e., they represent the true atmosphere state -- an assumption that can never be true, and will be discussed later). Then, the distances between the observation vectors and model state vectors will simply represent the observation errors. In other words, given an observation in the model domain and the model integration window, we can interpolate/derive the true state from the model at the observation time and location. The difference between the observation and the model will be the observation error. The difference is a “measure” of the magnitude of the observation error. Now let us assume that we have a perfect model and a perfect observation, but the model variables are grid-box average and the observation is a point observation. Then, the difference between the model and the observation will denote the representativeness error of the observation. Obviously, the

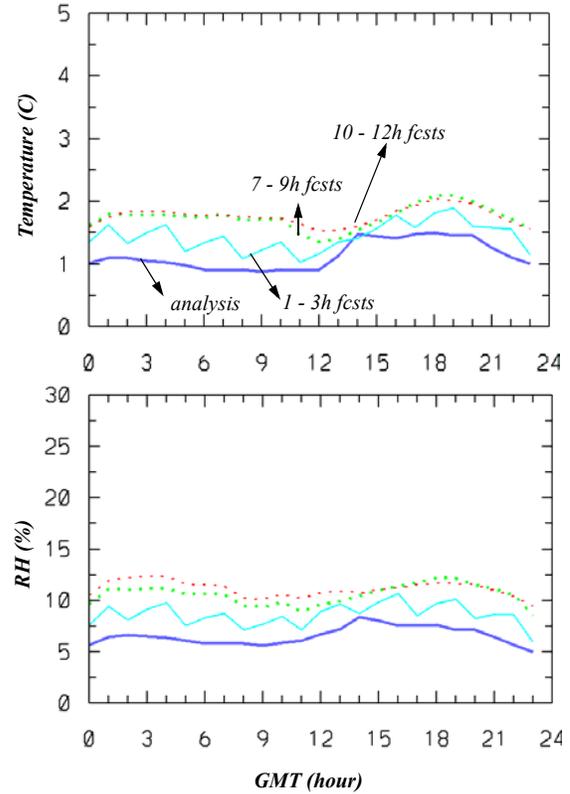


Fig.1 Average Mean Absolute Error (MAE) of surface temperature and relative humidity on Domain 2 (10 km grid) of the RTFDFA analyses and forecasts in August 2003, running at Aberdeen Proving Ground, Maryland.

representativeness error of an observation is in the model and observation difference.

It is beyond question that the RTFDFA analysis and forecasts have errors. Because of the existence of this model error, the difference between an observation and a model background results from both model errors and observation errors. Generally, we do not have enough information to separate the model and observation errors, based on a single case. However, the statistics of model errors and observation errors can be used. In the RTFDFA QC scheme, the statistics of model errors and observation errors are combined and used to scale the quality of observations according to the difference between the observations and the model values. It is important to recognize that, due to the existence of model errors, the observation quality should not be scaled linearly. The reason is that, if a model-observation difference falls within the model statistical error, we do not know if the difference is from the model error or the observation error. Nevertheless, observations may be unreliable when the model-observation difference is larger than the statistical error. As a matter of fact, the traditional procedure is based on this concept to cut off the “bad” observations. Accordingly, a Gaussian

distribution is used to scale the quality and generate generalized quality flags (Q_F):

$$Q_F = \exp[-(A_{\text{obs}} - A_{\text{model}})^2 / (\sigma^2 X^2)] \quad (1)$$

Where A_{obs} is an observation variable to be QC-ed, A_{model} is the same variable derived from the model, σ is the combined statistical model error and observation error ($\sigma^2 = \sigma_{\text{model}}^2 + \sigma_{\text{obs}}^2$) and X is a scaling parameter that specifies the weight distribution shape of the quality. X should be chosen based on data categories (measuring platforms and variables), the usage of the resultant quality flags, and experiences. Fig. 2 shows an example of the variation of the quality flag values for 2 choices of X , based on the model and observation error statistics used in the RTFDFA systems. A larger X will produce a larger (smaller) quality flag (confidence value) for an observation with smaller (larger) model-observation difference. It should be pointed out that the shape of the Gaussian curves also depend on the σ . Currently, σ is derived from model error statistics based on the GFS model, by using the "NCEP-Method" and the observation error statistics provided by NCEP. It should be pointed out that these σ values may not be ideal for the RTFDFA model and applications, and it is desirable to derive the σ value by comparing the RTFDFA analysis and 1- 3 hour forecasts with reliable observation sources (such as conventional radiosonde and METAR observations).

Finally, let us compare the new QC scheme with the traditional ones discussed in Section 2. The FDDA processes and dynamically incorporates observations into the full-physics model, and the observation information is spread in the model space through proper spatial and temporal weighting functions and through physical adjustment by the model dynamics. Therefore, compared with the traditional QC schemes, checking the observations against the RTFDFA analysis and/or 1 - 3 hour forecasts of the system implicitly takes account of 1) the "error-max-check" but with a better first guess; 2) the "buddy-check" but with better dynamical analysis buddies; 3) the "temporal consistency check", but with both model and observation time-continuity, and of course, 4) the "dynamical and physical check" under the strong constraint of the model physics and dynamics. However, unlike the "sharp cut-off" approach used in the traditional QC schemes (shown by the green line in Fig. 2), the new QC procedure "ramps down" the confidence/quality levels for an observation gradually from 1 to 0. More importantly, the new QC method scales the overall error of an

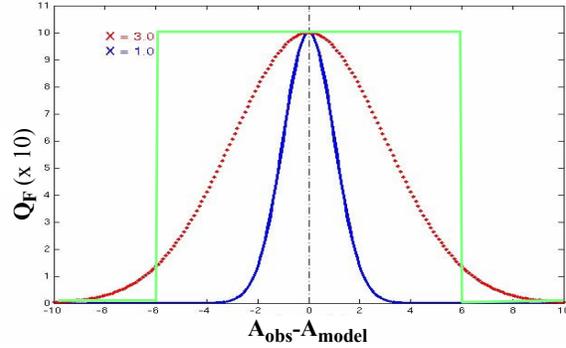


Fig.2 Variation of observation quality scaling function for two given X parameters. For references, the green line represents a result of the traditional "sharp cut-off" method.

observation, including the representativeness errors that varies with analysis grid sizes and weather conditions.

4. USE OF QUALITY FLAGS IN DATA ASSIMILATION

The "station-nudging" approach which was initially developed by Stauffer and Seaman (1994) is employed in the data assimilation of the NCAR/ATEC RTFDFA system. The system have been modified and improved in the ATEC RTFDFA applications (Cram et al. 2001, Liu et al. 2002a, 2002b, and Liu et al. 2003). The "station-nudging" approach dynamically relaxes model states toward the observation states by incorporating observations into the continuously running full-physics model. Observation innovations are added to the model momentum, mass and moisture equations at each time step, with proper space and time weighting functions that are centered at the observation time and location for each individual observation. The nudging terms (nudging tendencies) in the model equations can be written as:

$$\sum_{n=1}^N W_t W_h W_v G(A_{\text{obs}} - A_{\text{model}})$$

Here $A_{\text{obs}} - A_{\text{model}}$ is the observation innovation for each of the N observations, valid at the observation location; W_h , W_v and W_t are horizontal, vertical and temporal weighting functions, respectively; G is a nudging factor that specify the time period during which an observation innovation is active.

Unlike the 3DVAR and 3DPAS approaches, where statistical model (background) and observation errors are explicitly used in the analysis process to seek a "best-fit" analysis under simplified dynamic constraints, the 4-D nudging method implicitly reduces the total error of the model (backgrounds at each model time-step) and observations by maintaining model dynamical and physical balances. The

nudging approach uses $\mathbf{A}_{\text{obs}} - \mathbf{A}_{\text{model}}$ at the observation time and location to approximate the total error at the observation point. Nevertheless, the nudging does not provide information on the error structure (co-variances), and thus the spatial weighting functions (\mathbf{W}_h and \mathbf{W}_v) are specified by referring to past statistic results and adjustments with a large number of experiments.

Although the nudging approach can not separate the model (background) and observation error from the "total error", it does explicitly consider the weight of observations and the model. The nudging term (Eq. 2) depicts exactly how much the model state (background) should be corrected, with the weighting functions and the nudging factor defining how, and by how much, the model should be corrected. The products of nudging coefficient ($\mathbf{W}_h\mathbf{W}_v\mathbf{W}_t\mathbf{G}$) approximates the Kalman filter gain.

Nevertheless, since the weighting functions and the nudging factor are pre-defined constants, the nudging correction tendency varies proportionally to the model and observation differences. That is, the more the model diverges from the observation, the larger the correction that is imposed. This is perfectly fine if observations are perfect. However, observations have errors, and an observation possibly possesses larger errors when the total error ($\mathbf{A}_{\text{obs}} - \mathbf{A}_{\text{model}}$) is larger. With reduced confidence levels on those observations which differ more from the model, we obviously do not want to nudge the solution at a larger (or even the same) rate than (as) for other observations for which we have greater confidence. The quality flag obtained from the new RTFDDA QC scheme provides an extra weight that represents the confidence of an observation. It can be simply incorporated into the nudging term as a confidence weighting factor (Q_F):

$$\sum_{n=1}^N W_t W_h W_v G (\mathbf{A}_{\text{obs}} - \mathbf{A}_{\text{model}}) Q_F$$

5. EXPERIMENT RESULTS

To test the new QC scheme and the impacts of incorporating the quality flags in the RTFDDA analysis, two nested-grid, three-hourly-cycle RTFDDA simulations were carried out with the new and the traditional QC procedures. The same RTFDDA model configuration used during the OKC Joint Urban 2003 Atmospheric Dispersion Study field experiments (JU2003, Liu et al. 2004 - Paper 22.2 on this conference) was employed here. The model domains were defined over the Central Plains and were centered over Oklahoma City. A 5-day period, from 12Z Aug. 6 to 00Z Aug.

12, 2003, during which both clear-skies and severe convective weather occurred, was selected for the numerical experiments. The case was selected because of its abundant upper-air and surface data, including aircraft reports, wind profiler data, satellite winds and high-density and frequency observations from OK mesonet.

Table 1: Quality scaling parameter X, used in RTFDDA experiments

Variables	u/v	T	q
Non-radiosonde observation	1.3	1.5	1.3
Radiosondes	1.8	1.5	1.8

The traditional QC procedure used in the experiment includes "error-max-checks", "buddy-checks" and "consistency-checks". The maximum errors for the "error-max-check" and the "buddy-check" were set at 8 m/s for U and V wind components, 8 C for temperature, and 70% for relative humidity. In the experiment with the new QC procedure, the statistical model errors and observations errors, for radiosonde varying with height and METAR observations, were used to define the sigma in Eq (1). The X values in Eq. (1) are experimentally specified according to the properties of model-radiosonde difference. As can be seen in Table 1, the X values for radiosondes are a bit larger than those for other observations. This is because radiosondes are commonly used data sources and are considered more accurate and reliable than the other data sources. It should be pointed out that the new QC procedure can be used to estimate and monitor the overall quality of measurement platforms and, in turn, the statistical results can be used to refine the X values for the platforms. In the run with the new QC scheme, the quality flags were incorporated in the data analysis.

5.1 Data Quality-Control Results

Data QC procedures were run hourly for all observations valid within a time window from -30 to +30 minutes. Model (0 - 3 hour) forecasts valid at the exact hours (0 minute) are used as the first background. For simplicity, only QC results for the observations at 00Z and 12Z, when both synoptic and asynoptic data exist, are discussed. Fig.3 shows histograms of QC results for all surface and upper-air observations at 12Z, from the five-day simulation. It appears that the new QC scheme is able to reasonably ramp down the observation quality. About 70% of wind, 79% of temperature and 84% of relative humidity observations at the surface are designated as very good, with a quality flag of 9 or 10 (fig.3a). The quality

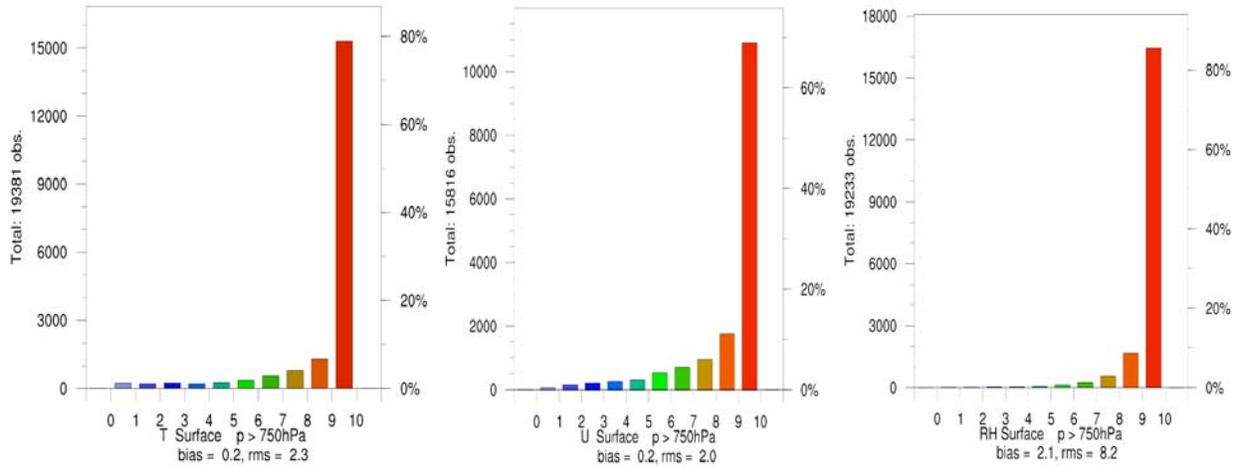


Fig.3a Quality spectra of the 2-m temperature, 10-m U wind component and surface relative humidity obtained with the RTFDDA QC scheme for all observations valid between 11:30 and 12:30 UTC of the 5 days experiment.

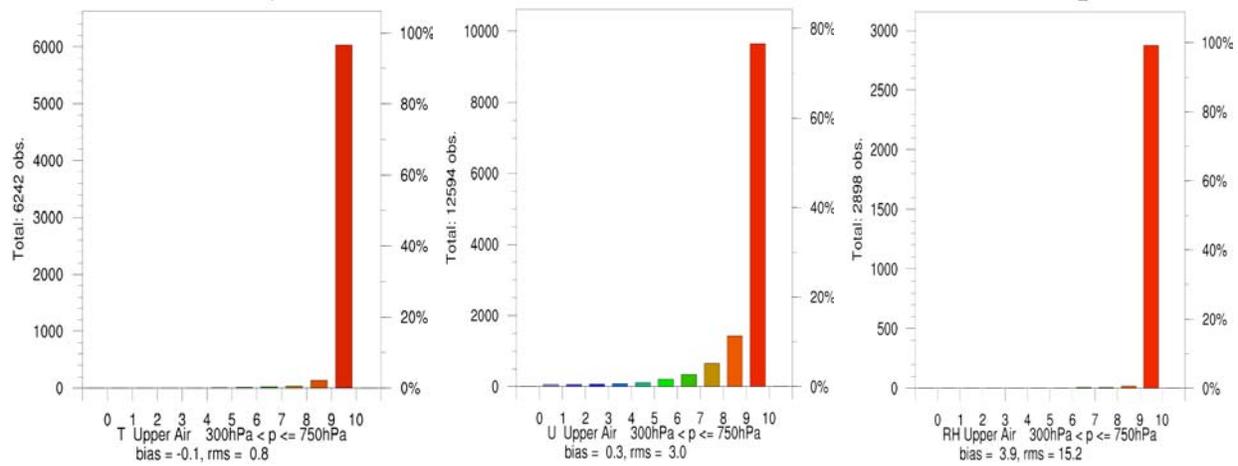


Fig.3b Same as Fig.3a, but for the middle troposphere (750 - 300 hPa) observation.

of upper-air data varies with altitude. Generally, observations in the mid-layer, between 750 hPa and 300 hPa, yielded larger quality flags (higher quality Fig.3b) than those in the layers below and above it. The quality spectra of the observations in the lower layers are similar to those at the surface. As was pointed out earlier, the scale parameter X can be adjusted based on the error properties of observations and the usage of the resultant quality flags. Refining X values will alter the shape of the quality spectra shown in Fig.3. For RTFDDA applications, the parameter settings allow the quality flags to be directly used as one of the weighting factors in the nudging term. By doing this, the result shown in Fig.3 indicates that the majority of the observations will be used with full confidence in the FDDA analysis.

Because observations are compared to a common baseline (background) during the QC process, it is possible, and of interest, to compare and monitor the quality of observations from various platforms with the new QC procedure. Fig.4 compares the observation-model wind scattering distribution at 00Z during the 5-day simulations for four major platforms: the conventional radiosondes (PILOT), NOAA/FSL NPN and CAP profilers (PROF), NOAA/FSL ACARS (AMDAR), and NOAA/NESDIS hourly GOES winds (SATWINDS), in the middle troposphere (750 - 300 hPa). Only the U component of the winds is shown in the plots (the V component is similar). Each cross represents an observation-model pair, with colors denoting the resultant quality values of the observations. Apparently, the PROF observations possess the smallest bias and dispersion from the

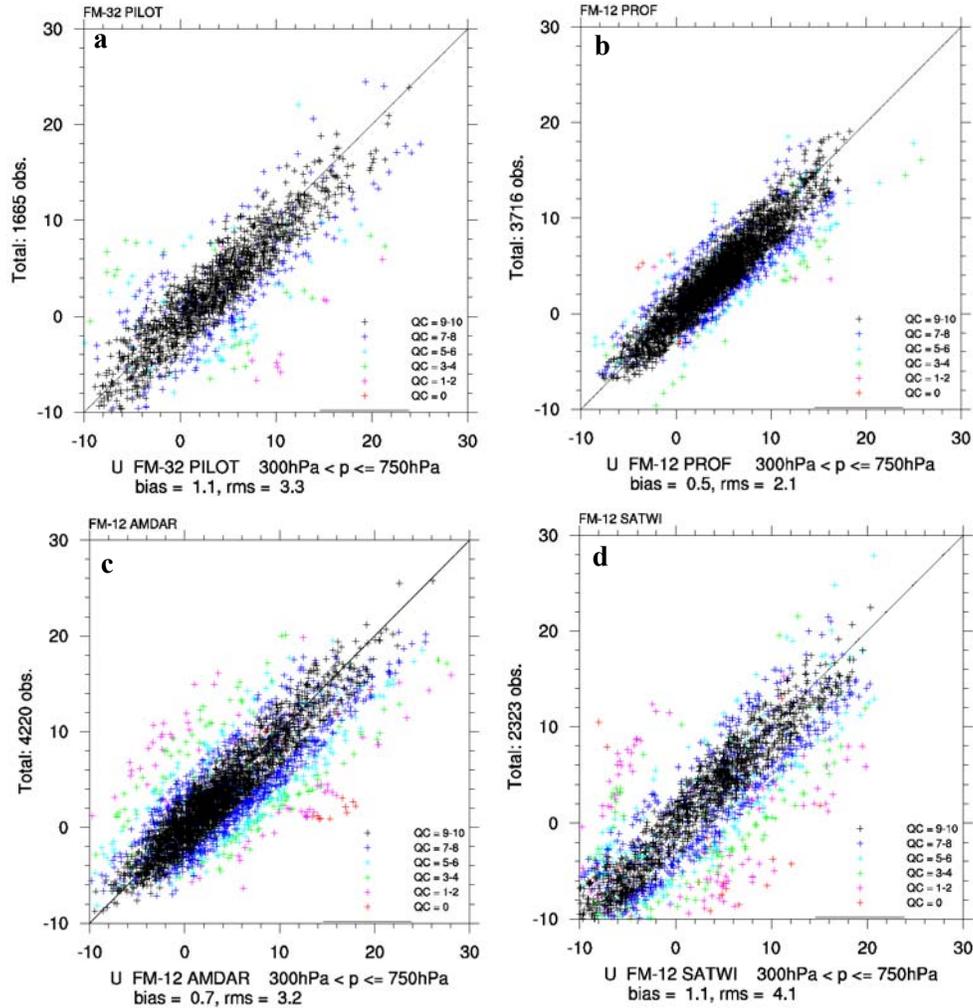


Fig.4 Observation (vertical coordinate) vs model (horizontal coordinate) scattering plots of the middle troposphere (750 - 300 hPa) U wind component of radiosonde (a), Profilers (b), ACARS (c) and GOES satellite winds, valid between 23:30 and 00:30 UTC, for the 5 day experiment. ($QC=10 \times Q_F$)

baseline (model) state, with a bias of 0.5 m/s and an RMS difference of 2.1 m/s. This is partially because the PROF data used here are hourly averages, which may have a smaller representativeness error than other point-observation platforms. The PILOT and AMDAR observations display very similar scatter properties, with RMS differences of 3.1 and 3.3 m/s respectively. Note that PILOT observations yielded overall larger quality flags because of the assignment of a larger X (more confidence) for the platform. A few clustered outliers with quality flag of 0 or 0.1 in AMDAR strongly suggest some errors associated with single flights. Finally, SATWIND observations are a bit more scattered than those from other platforms, with an RMS difference of 4.1 m/s.

It should be pointed out that, because the new QC method makes use of the statistical model and observation errors, one needs to be careful when examining the quality result of each

individual observation. The error of the model background may vary spatially and it may have bias. The same positive sign of the U wind bias from all platforms may suggest a systematic overestimate of U winds by the model in the layer, whereas the differences of the magnitudes of the bias between the platforms may suggest the existence of an observation bias of the platforms. Specifically, although SATWIND observations are generally less accurate due to the limitation of the retrieval schemes used. Also, these observations exist are mostly observed in regions where cloud and precipitation prevail where the model can be less accurate than in other regions. Thus the larger RMS difference between the SATWIND observation and the background may underestimate the quality of SATWIND observations.

The upper-air temperature used in the RTFDDA simulation mainly comes from the twice-daily radiosondes

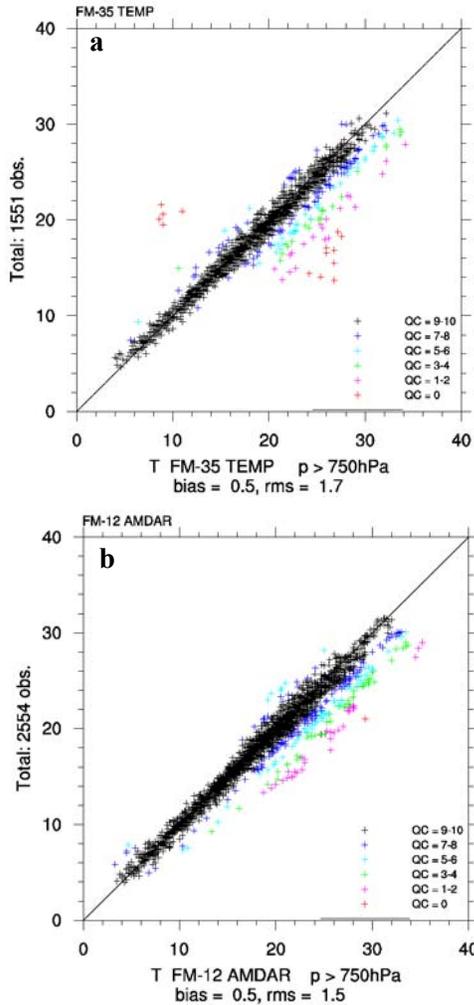


Fig.5 Same as Fig.4, but for temperature from radiosondes (a) and ACARS (b), with observations valid between 11:30 and 12:30 UTC in the lower troposphere (below 750 hPa).

(TEMP) and aircraft reports (ACARS/AMDAR). AMDAR observations are made at very irregular times and locations. Unlike winds, temperature observations from TEMP and AMDAR display much smaller dispersion between the model and observations. This is particularly true in the middle troposphere (between 750 hPa and 300 hPa, not shown). In the lower layer (below 750 hPa, shown in Fig. 5), some outliers can be seen in the boundary layer (i.e. the regions with warmer temperature in Fig.5). Both TEMP and AMDAR are 0.5 C colder than the model in this layer, and Fig.5 suggested that this systematic difference may be caused by a model warm bias.

5.2 Impact On RTFDFA Analyses and Forecasts

Verification of the RTFDFA analyses and forecasts were conducted hourly by interpolating the model to observation stations during the 5-day simulation. With 3-hourly cycling,

we can divide the analyses and forecasts in each cycle into forecast stages of three hour windows of forecast lengths, i.e. -3 - 0 hour final analyses, 1 - 3 hour forecasts, 4 - 6 hour forecasts... The same periods from all 8 cycles in a day will cover the full diurnal evolution for that forecast period.

Fig. 6 compares the average Mean Absolute Errors of the surface variables of the final analyses and 10-12 hour forecasts, run with the new and the old QC procedures, on the Domain 3. The results are mixed. The RTFDFA simulation with the new QC procedure appears to improve the surface analyses for all fields. It also produces better forecasts of the surface moisture, and some improvements to the surface temperature and wind direction forecasts during the daytime. Nevertheless, it slightly degrades the forecasts of wind speed and nighttime temperature.

The relatively small differences between the two runs may be because with the current X setting, about 80% of observations are classified as "near-perfect", the same as with the old-QC schemes, and 10% of the rest are weighted with a slightly lower weight that is still close to those with the old QC procedure. Large differences between the new and the old QC runs exist only for the other 10% of the "most questionable" observations. However, as pointed out earlier, efforts to properly define and scale observation-quality flags, and properly weight these flags in the data assimilation, are far from complete. Particularly, the statistical model error used in the current system is based on NCEP GFS model which is very different from the RTFDFA model. Benefit can be likely achieved by refining these procedures.

6. SUMMARY AND FUTURE WORK

NCAR/RAP and the Army ATEC program have been developing a multi-scale rapidly cycling real-time FDDA and forecasting system. The system produces spun-up FDDA analyses and forecasts in real-time, and incorporates observations from diverse measuring platforms. The 0 - 3 hour model forecasts from the system provide ideal three-dimensional weather states that can be used to examine observation quality, which may vary between observation platforms, and in space and time as well. In this paper, a simple and efficient observation data quality control (QC) procedure was constructed, based on the observation-model (the 0 - 3 hour model forecasts of the real-time RTFDFA system) difference and the statistics of model errors and observation errors. This QC procedure not only can objectively define the error tolerance criteria for elimination of most bad observations, but it can also estimate and assign a generalized quality flag to each individual observation, without regard to the platform by which it was observed. The quality flags, that represent varying degrees of confidence in

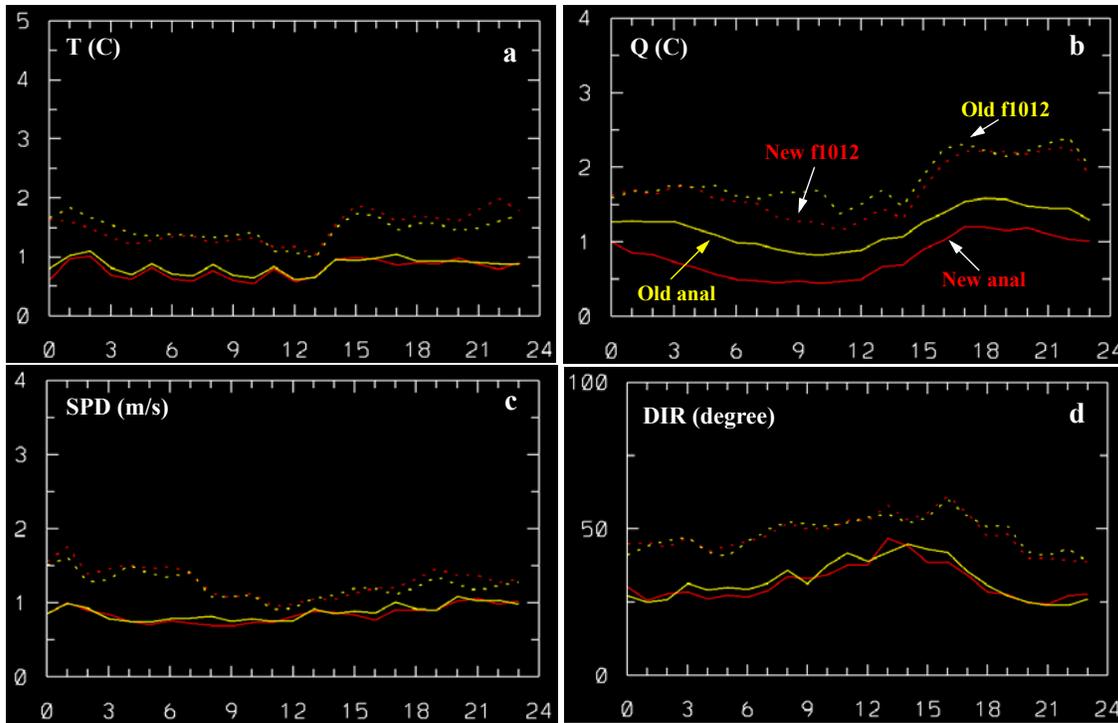


Fig.6 Average Mean Absolute Errors (MAE) of surface temperature (a), specific humidity (b), wind speed (c) and wind directions (d) of the RTFDAA final analyses (solid) and 10 - 12 h forecasts (dotted) on Domain 3, with the traditional (yellow) and new RTRFDAA QC (red) procedures.

observations (i.e. rough estimate of the probable errors), are determined according to the difference between the 0 - 3 hour model forecast and the observations, which are scaled by the statistical properties of the model-background and observation errors.

The quality flags of observations obtained from this QC procedure are incorporated into the four-dimensional data assimilation engine in the RTFDAA system. First, the quality flags are scaled to a value between 1 (the best quality) and 0 (the worst quality). Then, in the data analysis, each observation is weighted uniquely in the "station-nudging" term by its quality flag (or confidence factor). This approach allows the RTFDAA system to heavily weight the reliable observations and to lightly weight the less reliable observations.

Numerical experiments were conducted over a five-day period in the Central Plains with the model centered on Oklahoma City. Preliminary results show 1) a significant increase in the computation efficiency of the new QC scheme, because it skips many quality checking strategies used in the traditional QC procedure, and 2) reasonably good performance in discriminating outliers, and scaling the quality of observations. Statistical verification of two experiments conducted with the old (traditional) and the new RTFDAA QC procedure indicate some encouraging improvements in the

RTFDAA analyses and forecasts with the new RTFDAA QC procedure.

Finally, the new QC scheme is critically dependent on the accuracy of the model background. The better the model, the more accurate the QC, and, in turn, the better the model analysis is with the use of the QC information. Due to the existence of model errors, the QC scheme, should and can be, further refined according to the error properties of particular models and observations used in the application. This can be done using real-time runs or using a short history of the model results. For example, it is possible and desirable to estimate the systematic bias of both the MM5 model and the observation platforms, based on the statistics of the data quality generated by the RTFDAA system. The bias should be used to improve the QC procedure and the data analysis (Dee and Da Silva, 1998). On the other hand, some prior observation QC information, such as obtained by instrumentation calibration and from other QC schemes at data dissemination centers for certain platforms, can be collected and combined into the QC procedure to better quantify the final quality assignments. Finally, although, some data assimilation schemes such as 3DPAS and 3DVAR may not be able to make full use of the quality flags for each individual observation, the quality flags do provide a way to objectively define "cut-off" criteria to draw out "bad"

observations. Furthermore, since the model provides complete atmospheric states, the system may be extended to monitor the quality of observations of variables that are not directly forecasted by the model.

REFERENCE

- Cram, J.M., Y. Liu, S. Low-Nam, R-S. Sheu, L. Carson, C.A. Davis, T. Warner, J.F. Bowers, 2001: An operational mesoscale RT-FDDA analysis and forecasting system. *Preprints 18th WAF and 14th NWP Confs.*, Ft. Lauderdale, AMS, Boston, MA.
- Dee, D and A. M. da Silva, 1998: Data assimilation in the presence of forecast bias. *Quart. J. Roy. Meteor. Soc.*, **124**, 269-295.
- Hollingsworth A. and co-authors, 1986: Monitoring of observation and analysis quality by a data assimilation system. *Mon. Wea. Rev.*, **114**, 861-879.
- Liu, Y. and co-authors, 2002a: Performance and enhancements of the NCAR/A TEC mesoscale FDDA and forecasting system. *15th Conference on Numerical Weather Prediction*, San Antonio, Texas, 12-16 August, 2002, 399-402.
- Liu, Y. and co-authors, 2002b: Development and evaluation of a real-time FDDA and forecast system for the Year-2002 SLC Olympics. *12th PSU/NCAR Mesoscale model user's workshop*, 44-47, Boulder, Colorado, 24-25 June 2002.
- Liu, Y. and co-authors, 2003a: Improvements to surface flux computations in a non-local-mixing PBL scheme, and refinements to urban processes in the Noah land-surface model with the NCAR/A TEC real-time FDDA and forecast system, *16th Conference on Numerical Weather Prediction*, Seattle, Washington. Jan. 11-14, 2004.
- Stauffer, D.R., and N. L. Seaman, 1994: Multiscale four-dimensional data assimilation. *J. Appl. Meteor.*, **33**, 416-434.