

DTC Testing and Evaluation of Surface Data Impacts on the GSI Data Assimilation System

Kathryn M. Newman^{1,2}, C. Zhou^{1,2}, H. Shao^{1,2}, M. Hu^{1,3}, X.Y. Huang^{1,2}



¹Developmental Testbed Center (DTC)

²National Center for Atmospheric Research (NCAR)

³NOAA Global Systems Division (GSD)



Introduction

What is GSI?

The Gridpoint Statistical Interpolation (GSI) is a three dimensional variational (3D-Var) data assimilation system developed at NCEP/EMC, NOAA/GSD, NASA/GMAO, and NCAR/MMM. A community code version of the GSI system is also available and supported through the Developmental Testbed Center (DTC).

DTC T&E Objectives

- Perform GSI+WRF-ARW configuration runs to determine the capability and robustness of the GSI+ARW in regional applications and evaluate impact from a variety of existing and proposed new operational data types
- Provide rational basis for operational centers and the research community for advancements of NWP systems

Extended Tests

Experimental Design

The DTC set up an end-to-end system consisting of the WRF Pre-processing System (WRF, v3.2), GSI (Q1FY11), Boundary Update (UPDATE_BC) utility, Advanced Research WRF (ARW, v3.2), WRF Post-Processor (WPP, v3.1), and Model Evaluation Tools (MET, v2.0), for the GSI testbed.

Extended tests using this testbed were conducted using GSI Q1FY11 coupled with WRF-ARW v3.2. A one-month testing period was used from 15 August 2007 through 15 September 2007. The Air Force Weather Agency (AFWA) T8 domain was used for all extended tests (Fig1).

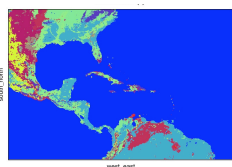


Figure 1: AFWA T8 Domain

The model was configured with a 15 km horizontal resolution, 57 vertical levels, and a 10 mb model top. The domain has 418 west-east and 280 north-south grid points. Configuration studies were performed, resulting in the DTC namelist closely representing that of the North American Model (NAM), with one exception being the qoption flag (qoption=1 was used in the DTC namelist). Forecasts were run for 48 hours at 00 and 12 UTC for verification purposes.

NCEP/EMC provided two types of Background Error (BE) statistics for the tests, the "global" BE statistics were calculated based on the GFS forecast differences and can be interpolated by the GSI code to any regional domain. Regional BEs were produced by the NAM model in a CONUS domain, reflecting the NAM specific background features and might introduce uncertainties the T8 domain study. Therefore, global BEs were used in all extended tests.

Four 1-month experiments were conducted for this study:

1. **GFSWRF**: Benchmark run. ARW runs initialized from the GFS analysis every 6 hours (cold-start) Three 6-hour full cycling runs using GSI included:
2. **CYC_CONV_default**: GSI and ARW assimilating conventional PREPBUFR data
3. **CYC_CONV_allobs**: conventional PREPBUFR data were assimilated with increased surface observations
4. **CYC_CONV_nosfc**: No Surface PREPBUFR data were assimilated

The 6-hour full cycling runs were initialized with the 6-hour WRF ARW forecast from the previous cycle to conduct impact studies. For all runs, the lateral boundary conditions (LBC) came from the GFS forecast.

All experiment results were verified using MET v2.0 against GDAS conventional PREPBUFR data. Statistics were computed at 17 vertical levels up to 10 mb. Statistical significance was tested at the 95% confidence interval.

Acknowledgments

This work was funded through the DTC by the Air Force Weather Agency (AFWA). NCAR is sponsored by the National Science Foundation (NSF).

Assimilation Results

Data Impact : GFSWRF vs. CYC_CONV_default

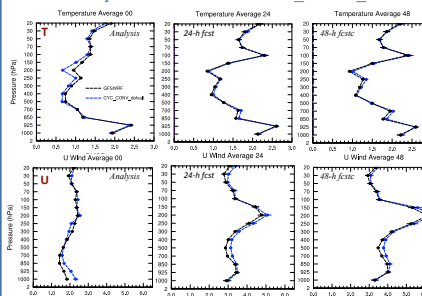


Figure 2: Vertical profiles for the analysis (left), 24-hour forecast (middle), and 48-hour forecast (right) for the temperature (upper) and u-component wind (lower) for GFSWRF (black) and CYC_CONV_default (blue).

- The analysis and forecasts generated from CYC_CONV_default were first compared with those from GFSWRF to ensure the cycling runs were correctly conducted.
- CYC_CONV_default only assimilated PREPBUFR conventional observations, where the initial conditions for the GFSWRF run are from the GFS analysis with all operational data types assimilated.
- Monthly scores are over long time scales and large-scale dynamics
- Statistically significant (SS) differences in the initialization of the GSI run compared to the benchmark GFS run. Forecast times show neutral to slightly negative differences favoring GFSWRF

Data coverage: PREPBUFR conventional over T8 domain

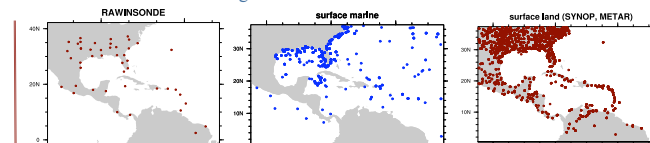


Figure 3: Data coverage of PREPBUFR conventional data over T8 domain. Includes rawinsonde (120), surface marine (180,182), surface land (181,187)

Surface Assimilation Impact

In order to evaluate the impact of surface data on the GSI assimilation system, a series of surface impact runs were performed. CYC_CONV_default assimilated observations with default GFS quality marks for the GDAS data. These QC marks resulted in most of the surface data being rejected during the GSI minimization. CYC_CONV_allobs loosened the QC mark threshold values for the surface data assimilation to match the data types allow in the NAM. This process may introduce some bad observations to the analysis because the NCEP NAM QC procedure was essentially skipped, however the purpose of this run was for evaluating the impact of surface observations and any potential additional data. A third run, CYC_CONV_nosfc was conducted which removed all PREPBUFR surface observations from the assimilation.

Table 1 shows the QC mark value for each corresponding PREPBUFR report type. All runs followed NAM configurations, while changing the 'qmark' threshold altered the amount of surface data assimilated. CYC_CONV_default qmark was set to 2.0 (allowing obs in green), CYC_CONV_allobs qmark was altered to 9.00 (allowing obs in green and red)

Table 1: QC mark value for corresponding PREPBUFR report type. Green values used in CYC_CONV_default run, red and green values used in CYC_CONV_allobs, and all observations in table were not used in CYC_CONV_nosfc.

	ADPSFC 181	ADPSFC/ SFCSPH 183	ADPSFC 187		ADPSFC 281	SPSSMI 283	ADPSFC/ SFCSPH 284	QKSWND 285	ADPSFC 287
Ps	2.00	9.00	6.00	U V	9.00	2.00	9.00	2.00	9.00
Q	9.00	9.00	9.00						
T	8.00	8.00	8.00						

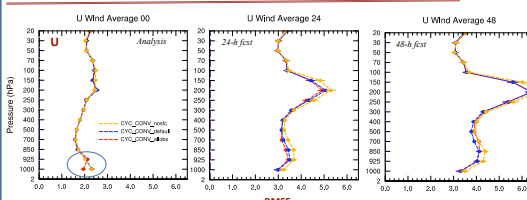


Figure 4: Vertical profiles for the analysis (left), 24-hour forecast (middle), and 48-hour forecast (right) for the u-component of the wind field for CYC_CONV_default (blue), CYC_CONV_allobs (red) and CYC_CONV_nosfc (yellow). Verification against sounding data.

The impact of surface observations at the analysis time is evident (Fig 5), with little impact by the 48 hour forecast. Assimilation of surface data reduced the wind bias through the 48 hour forecast. This positive impact (Fig 4) has propagated from the surface at the analysis time into the upper levels at the 24-hour and 48-hour forecast times.

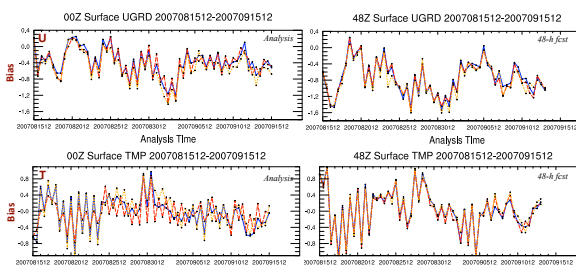


Figure 5: Time series of surface temperature (lower) and u-component wind (upper) bias for analysis (left) and 48-hour forecast (right) for CYC_CONV_default (blue), CYC_CONV_allobs (red), and CYC_CONV_nosfc (yellow). Verification against sounding data.

Conclusions

- The surface impact study showed positive impacts from surface data assimilation on the near surface analysis and upper air at the forecast times.
- Simply adding more surface observations will not increase forecast skill, however, removing all surface observations will provide a SS negative impact compared to using the default number of surface observations following the GFS QC marks.
- Further studies need to be conducted to determine how to increase surface data usage during assimilation with resulting forecast skill

Author contact : knewman@ucar.edu