Climatologies Based on the Weather Research and Forecast (WRF) Model

Francois Vandenberghe, Mike Barlage, and Scott Swerdlin
National Center for Atmospheric Research (NCAR) and the STAR Institute, Boulder, CO
{vandenb, barlage, swerdlin}@starinst.org

Judith Gardiner, Ashok Krishnamurthy, and Alan Chalker
Ohio Supercomputer Center, Columbus, OH
{judithg, ashok, alanc}@osc.edu

Because tests at Army Test and Evaluation Command (ATEC) ranges can often only be conducted under specific atmospheric conditions, climatologies of each range could be a very useful tool for long-range test planning. Such datasets can provide the probability density function of near-surface wind speed, percent cloud cover, temperature, precipitation, turbulence intensity, upper-atmospheric wind speed, etc., as a function of the month of the year, time of day, and location on the range. The STAR Institute with guidance from the Ohio Supercomputer Center is working on porting the Climate Four Dimensional Data Assimilation (CFDDA) technology developed by the National Center for Atmospheric Research to version 3 of the Weather and Research Forecast (WRF) model for use on systems within the DoD High Performance Computing Modernization Program. The fully parallelized WRF-CFDDA system uses data-assimilation by Newtonian relaxation, to generate a regional re-analysis that is consistent with both observations and model dynamics for every hour of the past 20 years, at fine scales (3 km grid) over the Dugway Proving Ground range.

1. Introduction

Over the past years, the National Center for Atmospheric Research (NCAR) has developed the Climatology Four-Dimensional Data Assimilation system weather system (CFDDA). CFDDA has the capability to generate hourly atmospheric analysis at a horizontal resolution down to the kilometer for any meteorological situation or time period in the past 30 years. The previously developed system consists of the following model and observation input data:

1. The 5th generation of the NCAR/Penn State Mesoscale Model (MM5) and its Four Dimensional Data Assimilation (FDDA) functions
2. 40 years of global analysis generated at NCAR with the National Centers for Environmental Prediction (NCEP) model (NNRP)
3. 30 years of meteorological observations from the Automatic Data Processing historical repository maintained at NCAR
4. 28 years of high resolution of Sea Surface Temperature reanalyses from NCEP
5. 20 years of the NASA's Global Land Data Assimilation Surface model output land and soil data

NCAR MM5 CFDDA runs on Linux clusters and has already been applied in various Defense applications for several sponsors (NGIC, DTRA, and AMREOC). The system could also prove to be a very valuable mission planning tool for the Army Test and Evaluation Command (ATEC) ranges. For example, the 20–40 year climatologies created by CFDDA can provide the detailed probability density function (probability of different values) of near-surface wind speed, percent cloud cover, temperature, precipitation, turbulence intensity, upper-atmospheric wind speed, etc., as a function of the month of the year, time of day, at any location on the range. An analysis of the climatology could therefore help the ranges to define the season, time, and location on the range where required test conditions are most likely to be met—and long-range scheduling and planning for tests can be undertaken accordingly. The ATEC ranges are distributed over the CONUS and Alaska. Generating 20 years model gridded climatologies for each of these ranges is an exceptional computational challenge that requires resources not normally available to academic institutions.

STAR Institute is working with NCAR under the guidance of the Ohio Supercomputer Center on developing a new CFDDA capability based on the third version of the Weather and Research Forecast (WRF) model. This new version of the WRF model includes several important features, such as Raleigh damping near the top of the model, which are highly relevant for range meteorology. The code is fully parallelized making WRF very suitable for implementation in the High Performance Computing Modernization Program (HPCMP) distributed memory computing environment.

The WRF code with its preprocessor, postprocessor and peripherals has been ported to the Air Force Research Laboratory Department of Defense (DoD) Supercomputing Resource Center (AFRL DSRC) Linux cluster Falcon. The input data, about 240GB compressed, have been uploaded to the archive. A verification and validation of the porting of the code has been conducted and a statistical post-processing of the validation results has been applied. The WRF model and the configuration adopted for the climatological study is described in Section 2. The verification results are given in Section 3 and some of preliminary statistical products are presented in Section 4. Section 5 concludes the report with some considerations on the statistical and numerical aspects of the climatologies that are created.

2. WRF Configuration

The month of September over the Dugway Proving Ground (DPG) range was selected to conduct the verification. This choice was primarily motivated by the high quality meteorological observations available at this location at this time of the year after collection during the field campaign FFT07. During the FFT07 campaign, the surface local measurements routinely collected at the ranges through its mesonet SAMS network were transmitted in real-time to the MADIS database, and can, therefore, be found in the NCAR historical observational database (ADP). The month of September 2006 was chosen to ensure that range surface reports are only ingested into the WRF model when the SAMS database is used, and gives us more control on the data withholding experiments. The Dugway WRF Real Time Time FDDA real-time set-up (Liu et al., 2008) was used, except that the smallest and fourth domain at 1.1km grid increment is not activated in the C-FDDA system. This leaves three embedded domains with a resolution of 30/10/3.3km, respectively. The domain configuration is illustrated in Figure 1. In such a configuration each domain provides boundary forcing to the smaller domain it encompasses. There are 37 vertical levels, and the first vertical level at the center of the domain, where most of the local surface stations are located, is approximately 2 meters above the ground. In the verification, however, the model diagnostic 2-meter temperature and humidity and the 10-meter winds, and not the lowest vertical level, were used.

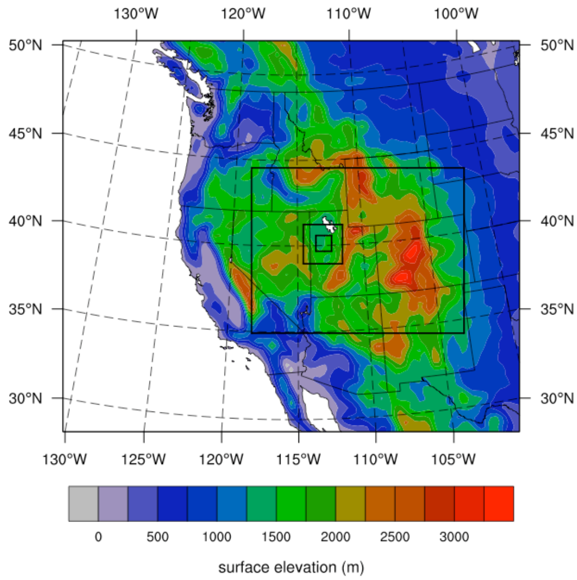

Figure 1. Terrain height and WRF Domain configuration over Dugway Proving Ground. Grid increment for Domain 1 (windows), 2 (largest rectangle), 3 (medium rectangle) and 4 (smallest rectangle) is, respectively, 30km, 10km, 3.3km and 1.1km. Domain 4 is not activated for climatological runs.

Initial conditions and lateral boundary conditions for the largest domain are taken from the 6-hourly 2.5-degree analyses of the NCEP/NCAR Reanalysis project data files archived at NCAR (http://dss.ucar.edu/datasets/ds091.0/). The model is reinitialized every 8.5 days (i.e., cold started at 12z the last day of the preceding month and then on the 8th, 16th, and 24th of the month) and output from the first 12 hours after cold-start is thrown away. Observations are ingested at the time they were recorded. The model integration takes 45 min. wall clock time per day of simulation on one of the 1,024 nodes of the AFRL DSRC machine "Falcon". The WRF physical parameterizations adopted for the runs are listed under Table 1.

Table 1. WRF parametrizations used for the simulations

Physical Parameters	Scheme
Cumulus Parametrization	Kain Frictch
Microphysics	Lin et al.
Long Wave Radiation	RRTM
Shortwave Radiation	Dudhia
Boundary Layer	YSU
Surface Layer	Monin Obukov
Land Surface	Noah

3. Verification and Validation

The verification and validation of a reanalysis mesoscale system poses an interesting question regarding the treatment of independent observations. Ideally, one would like to evaluate the analysis system at its best performance, that is, when the system uses all the available information, which would leave no independent observational data set. Withholding some of the observations for verification will lead to the validation of a sub-optimal system. Still, keeping this in mind, verification against independent data appeared to us to be an important step of the validation process. And since the project is focusing on the ATEC ranges, the local surface observations collected every day at the ranges offers an ideal dataset for this type of verification process. Those observations are issued from automatic stations, referred as SAMS, distributed over the ranges. SAMS report 2-meter temperature and humidity and 10-meter winds with high temporal frequency, every 5 min. for certain ranges. For a given range, the numbers of SAMS stations and their geographical locations have changed over the years, but NCAR has consolidated a database of more than 10 years of SAMS records that can provide useful information on the surface weather at the ranges.

SAMS reports have been transmitted to the National Oceanic & Atmospheric Administration (NOAA) data collection network through the MADIS database since 2007, but for all prior years, SAMS data are not present in the NCAR ADP database used by CFDDA. They are available, however, from NCAR and assimilated during generation of the ranges climatologies. The verification statistics of WRF CFDDA output against SAMS surface reports for the month of September 2006 over Dugway Proving Ground are presented below.

There were 19 SAMS stations reporting during the month of September 2006. On average, one report every 5 min. was collected in the NCAR SAMS database. After application of the RT-FDDA quality control procedure described by Liu *et al.* 2004, about 3800 reports were available each day for verification over the full month. For each report, the 2-meter temperature and humidity and the 10-meter wind components gridded field from the hourly WRF output files nearest to the reporting time were horizontally interpolated to the SAMS locations. The bias and root mean square difference between the observed and the model interpolated values were calculated for each report and aggregated by hour of the day for the full month. Tables 2, 3, 4, and 5 present those statistics for wind components, temperature and relative humidity at 00z (18MDT) over domain 3 for three different runs:

a) NOBS: no observations assimilated
b) ADP: ADP only observations assimilated
c) ADP+SAMS ADP and SAMS observations assimilated.

Table 2. Wind U component aggregated verification statistics at 00z against SAMS data of WRF domain 3 for the month of September 2006

U-wind	NOBS	ADP	ADP+SAMS
Bias in m/s	0.5	0.3	0.5
Rms in m/s	3.0	2.9	3.3

Table 3. Wind V component aggregated verification statistics at 00z against SAMS data of WRF domain 3 for the month of September 2006

V-wind	NOBS	ADP	ADP+SAMS
Bias in m/s	-1.2	-1.0	-0.6
Rms in m/s	2.8	2.9	3.0

Table 4. Temperature aggregated verification statistics at 00z against SAMS data of WRF domain 3 for the month of September 2006

Temperature	NOBS	ADP	ADP+SAMS
Bias in K	-1.2	-0.3	-0.1
Rms in K	3.2	2.2	2.9

Table 5. Relative Humidity aggregated verification statistics at 00z against SAMS data of WRF domain 3 for the month of September 2006

Relative Humidity	NOBS	ADP	ADP+SAMS
Bias in %	7.1	5.0	2.1
Rms in %	12.8	10.0	8.1

The statistics averaged across the domains are in agreement with the performances of current mesoscale models. It is striking, but not unexpected, that the assimilation of observations tends to slightly degrade the accuracy of the winds analysis. As it was noted in Rife and Davis 2005, the DPG range has complex terrain with sub-kilometer features that cannot be completely represented by a model with 3.3km grid increment. As a result, the model has difficulties reproducing all local effects (e.g., canyon, canopy, etc.) that are immediately around the recording station. These sub-grid properties particularly affect winds. For these stations, it is known (Davis et al., 2005) that a model with coarser resolution, which produces smooth fields, could, on average, produce better statistics than a higher resolution model. Figure 2 depicts graphically the impact of the assimilation.

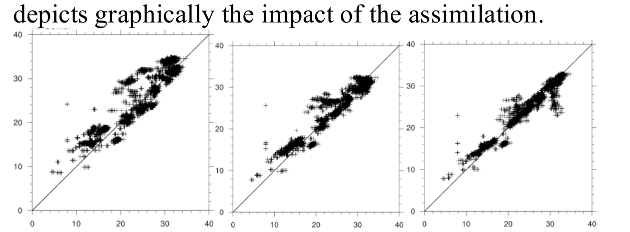

Figure 2: Distribution of the SAMS temperature pairs (observation on the x-axis/model interpolated value on the y-axis) for domain 3 for NOBS (left), ADP (center) and ADP+SAMS (right) WRF simulations.

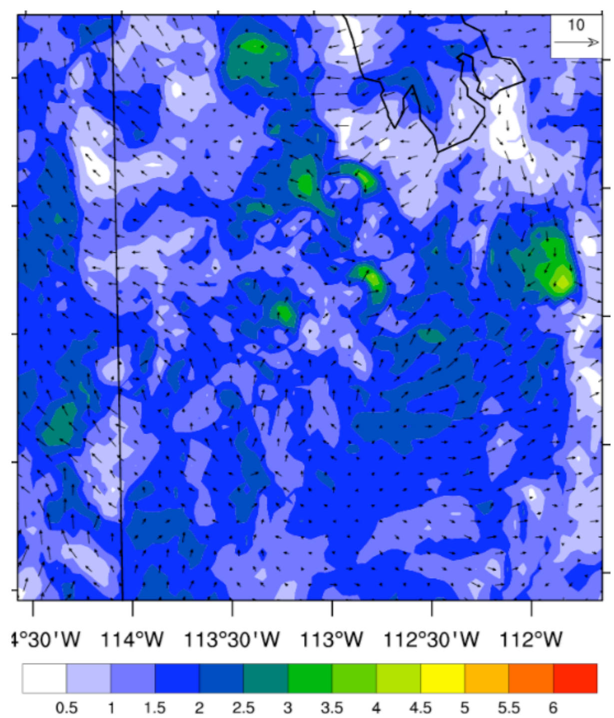

Figure 3. 10m climatic mean wind (vectors) and standard deviation (color) for September at 00z over DPG test range. Years 2000–2006 were used.

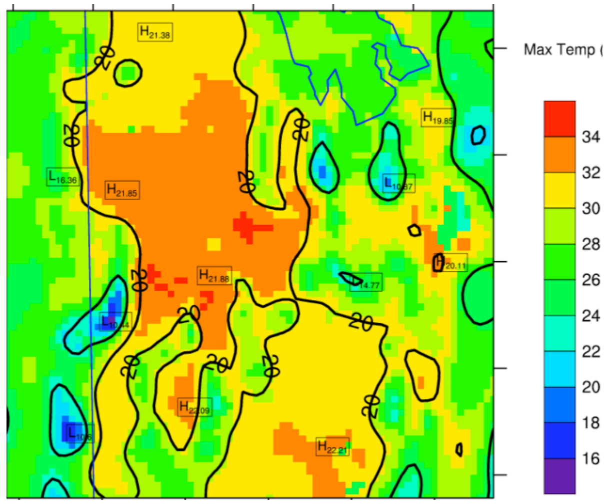

Figure 4: 2m temperature minima (contour) and maxima (color) at 00z over DPG test range. Years 2000–2006 were used.

4. Preliminary Statistical Results

A statistical post-processing was applied to hourly WRF output files for the month of September 2000-2006. For each hour of the day, the 30 days × 7 years = 210 files were averaged, creating a mean value for each meteorological field at each point of the model grid. The standard deviation was then calculated. Figure 3 shows the mean wind vectors at 10m with the standard deviation overlaid in color at 00z. The hourly climatic extremes, whose values are used in the testing decision making process, are also calculated from the collection of files. The locations of high variability coincide with the location of elevated terrain objects indicating that, although not perfectly accurate, the 3.3km domain is able to capture small-scale processes such as canyon channeling and upslope-downslope winds. Figure 3 shows the minima in contour line and the maxima in color of the 2m temperature.

5. Conclusions

A capability to efficiently generate years of high-resolution climatologies at the ATEC ranges with the WRF model has been developed and ported onto HPCMP super-computers. This capability is based on NCAR FDDA technology and uses various public global databases (NCAR/NCEP Reanalysis, NCAR ADP, NCEP OI SST, and NASA GLDAS) as input data. A month of data for September 2006 has been generated over the DPG range and verification statistics against independent local and global data have been presented. The results indicate that the accuracy obtained with this climatological version of the WRF FDDA system is similar to the accuracy obtained at the ranges by its real-time counterpart WRF RT-(?)FDDA and an earlier version of the CFDDA system based on MM5.

The HPCMP AFRL DSRC Linux cluster "Falcon" was used in this study. The month of simulation was decomposed into four 8.5 day-long independent jobs which were conducted in parallel in about 21 hours wall-clock time using a total of 8 processors of the 2,048 of the machine. Increased speed-up can be achieved with more computational resources, i.e., more than four nodes per month of simulation, but this acceleration will have to be achieved through WRF (geographical) domain decomposition. Figure 5 shows the shape of the WRF speed-up for an elemental job (8.5 days) as function of the number of processors used for domain decomposition, based on series of benchmarks conducted on "Falcon". The 4-2-processor configuration was used in this study.

Because the speed-up does not increase linearly with the number of processors, for a given number of processors, it is more efficient to run as many 2-processor 8.5 day elemental jobs as possible. For example, if 200 hundred processors are made available, the 30-year climatology can potentially be generated in 13 days under this configuration, while it could take up to 32 days if 32 processors are allocated to the elemental 8.5 day job. The final duration to produce the climatology will depend on the number of processors available at a given time on the machine.

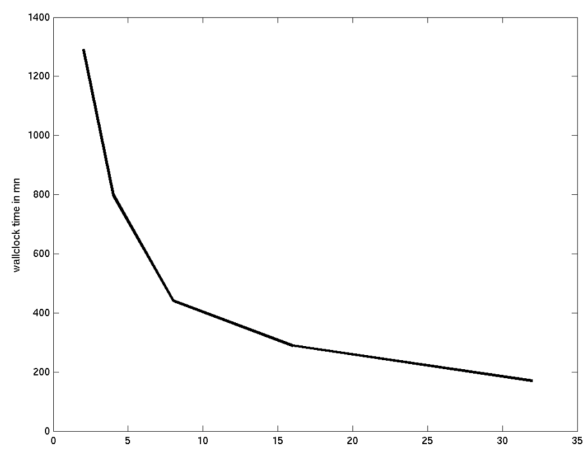

Figure 5. WRF speed-up as function of processors for the WRF configuration of Figure 1

Acknowledgement

This publication was made possible through support provided by DoD HPCMP PET activities through a contract with Mississippi State University The opinions expressed herein are those of the author(s) and do not

necessarily reflect the views of the DoD or Mississippi State University.

References

Davis, C.A., B.G. Brown, D.L. Rife, and R. Bullock, "Object-based verification approaches for NWP." *21. Conf. on Weather Analysis and Forecasting/17. Conf. on Numerical Weather Prediction*, Washington, DC (USA), 31 Jul–5 Aug 2005.

Laprise R. et al., "Challenging some tenets of Regional Climate Modelling." *Meteorol Atmos Phys*, 100, pp. 3–22, 2007.

Liu Y. et al., "The Operational Mesogamma-Scale Analysis and Forecast System of the U.S. Army Test and Evaluation Command. Part I: Overview of the Modeling System, the Forecast Products, and How the Products Are Used." *J. Appl. Meteorol. Climatol*. Vol. 47, no. 4, pp. 1077–1092, 2008.

Liu, Yubao;, F. Vandenberghe, S. Low-Nam, T. Warner, and S. Swerdlin, "Observation-quality estimation and its application in NCAR/ATEC real-time FDDA and forecast (RTFDDA) system." *AMS Symp. on Forecasting the Weather and Climate of the Atmosphere and Ocean, Seattle*, WA (USA), 10–16 Jan 2004.

Rife, D.L. and C.A. Davis, CA, "Verification of Temporal Variations in Mesoscale Numerical Wind Forecasts." *Mon. Weather Rev*., Vol. 133, no. 11, pp. 3368–3381, 2005.