Final Report

To

Environmental Protection Agency (EPA)

Purchase order EP08D000663

Reporting Period: 1 September 2008 – 31 January 2010

'Stable Boundary Layers Representation in Meteorological Models in Extremely Cold Wintertime Conditions'

Dr. Brian J. Gaudet, co-PI

Dr. David R. Stauffer, co-PI

The Pennsylvania State University

Dept. of Meteorology

University Park, PA 16802

bjg20@met.psu.edu

12 May 2010

EXECUTIVE SUMMARY

This final report describes work performed by Penn State for the EPA-funded Purchase Order EP08D000663 titled 'Stable Boundary Layers Representation in Meteorological Models in Extremely Cold Wintertime Conditions'. The purpose of the project was to develop, adapt, and test a methodology for stable boundary layer representation (initial onset, space/time evolution, dissipation) in three-dimensional numerical models, with a specific focus on the dark, extremely cold environments such as those in the winter in the Fairbanks, AK region. A particular concern is the frequent occurrence of very high fine particular matter (PM2.5) concentrations within the stable boundary layers that form in these conditions.

Ten tasks were defined in the Statement of Work (SOW) for this project. A summary of these tasks and a brief overview of the work completed can be found in the Appendix to this report. Two twenty-day episodes were selected from the 2007-2008 winter season to study periods of extremely cold temperatures and high PM2.5 concentrations and to evaluate model performance: one in near total darkness (14 Dec 2007 – 03 Jan 2008), and the other in partial sunlight (23 Jan 2008 – 12 Feb 2008). One baseline physics configuration and three physics sensitivity experiments were performed for each episode. The physics sensitivity experiments were used to assess the impact of different planetary boundary layer (PBL) parameterizations, land surface models, and atmospheric radiation schemes on the simulations. Each simulation used three nested grids: Grid 1 (12-km horizontal grid spacing) and Grid 2 (4-km) utilized the multiscale multigrid data assimilation strategy of Stauffer and Seaman (1994) in order to ensure the model and observations remain close over the extended duration of the simulations, and Grid 3 (1.3-km) did not use any direct data assimilation, and so was best-suited for quantifying the physics sensitivity. Grid 3, which is centered over the Fairbanks region, also possesses sufficient horizontal resolution to be used by the EPA as meteorological input to chemical and air transport and dispersion models. From the different physics packages one was to be recommended to the EPA for further mesoscale modeling of the region.

The major findings and impacts of this project are as follows:

- The use of the three-grid configuration with a multiscale, multigrid four-dimensional data assimilation (FDDA) strategy on the outer two grids and no direct FDDA on Grid 3 consistently produced qualitatively plausible atmospheric fields throughout the variety of meteorological conditions found in the episodes, despite the relatively sparse data density. Quantitatively, the multiscale, multigrid FDDA strategy led to improved root-mean-square-error (RMSE) scores for both wind and temperature on all grids. The FDDA on the outer domains had the desired effect of improving the simulations of Grid 3 without FDDA and used for physics sensitivity tests, by providing improved lateral boundary conditions.

- The best RMSE scores for the combination of both surface and sounding data required modification of the default FDDA procedure. These modifications included applying surface wind observational data to the third model vertical level instead of the lowest model level because wind observations are normally taken at a height of 10 m which is the height of the third level in the high vertical resolution configuration used here. The influence of surface observations was also restricted to approximately the lowest 100 m, instead of the top of the PBL, because the model-predicted PBL height in these simulations, based on the turbulent kinetic energy profile, was often found to be 1 km or higher. This correction applied the surface innovation (observation minus model value) in these predominantly stable boundary layers over a much shallower layer and produced improved statistical results in the lower troposphere.

- All model physics combinations tended to have a positive temperature bias on Grid 3, especially during the most extremely cold periods. All of the physics sensitivity tests tended to reduce the warm bias in comparison with the selected baseline physics package.

- Switching from the RRTM longwave / Dudhia shortwave radiation package to the RRTMG longwave and shortwave radiation package led to significantly reduced warm biases and better RMSE statistics. RRTMG was then used in all future physics sensitivity tests. The reduced warm bias seemed to be due to the longwave component, both because of direct examination of surface fluxes in the partial sunlight case, and due to the fact that the difference was more pronounced in the near total darkness episode.

- The simulation with the Rapid Update Cycle (RUC) land surface model, the Mellor-Yamada-Janjić (MYJ) PBL model, and the RRTMG radiation package was the coldest of the four physics suites tested, and had the lowest positive temperature bias and best statistics during those periods when the temperature was coldest. It was thus selected as the physics configuration of choice, since the coldest temperature conditions are those with the potential for the highest PM2.5 concentrations. However, there were periods in each episode, generally when the temperature was steadily decreasing in advance of an extremely cold period, during which the models had a cold bias. During these periods the RUC/MYJ/RRTMG configuration would usually be even colder and thus have worse magnitude temperature biases and RMSE scores. Thus, while this configuration was recommended, we also strongly recommended that the final fine-scale atmospheric data (i.e., from Grid 3) to be provided to EPA should come from an additional simulation in which FDDA is performed directly on Grid 3, in order to reduce some of this error.

- Wind component and wind speed statistics generally showed much less variability among the model physics sensitivity experiments than that seen for temperature. The MYJ/RUC/RRTMG (MRR) configuration usually produced slightly better wind statistics than the other configurations.

- Use of obs nudging for temperature and humidity (and not surface wind) on Grid 3 produced large improvements in the mass fields as expected, and also improvements in the wind fields

above the surface.   Results were very encouraging and suggested that a smaller (larger) time window should be used for the surface (above-surface) data assimilation.   This capability present in the Penn State MM5 FDDA system has been added to the new-release version of WRF.

- In addition to this final report, deliverables to the EPA will include the full three-dimensional output at relatively fine temporal resolution (every 1 hour for Grid 1; every 12 minutes for Grids 2 and 3) for the final Grid 3 nudging simulation as well as all the baseline and physics sensitivity simulations.   Model namelists, initialization files, and modifications to the model source code will also be provided.

- The development and refinement of WRF FDDA capabilities and supporting software, including the surface analysis nudging, observation nudging and the OBSGRID objective analysis and obs-nudging pre-processing code, occurred concurrently with this project.   This separate development effort led by PI Dave Stauffer and funded by the Defense Threat Reduction Agency (DTRA) allowed us rapid access to the most recent and robust versions of the WRF FDDA code, and this greatly benefited this project.

- The results of the default FDDA procedures not performing well in this high vertical resolution modeling study of stable boundary layer environments motivated an additional FDDA code development effort to make the vertical influence functions of surface observations within the FDDA be a function of stability regime type, as well as to provide the user with greater flexibility in specifying the vertical influence functions.   These modifications were not finalized in time to be used for this project but are scheduled to appear in the next official release of the WRF model.

- An extended abstract and oral presentation were made at the 13th Conference on Mesoscale Processes (Gaudet et al. 2009), and a manuscript based on the project is in preparation.

- Since the first draft of the final report, the Grid 3 FDDA design and simulations have been completed for both twenty-day episodes.   The results showed that the use of obs nudging for temperature and humidity (but not surface wind) on Grid 3 produced large improvements in the mass fields (as expected), and also improvements in the wind fields above the surface.   Results were very encouraging and suggested that a smaller (larger) time window should be used for the surface (above-surface) data assimilation.   This capability present in the Penn State MM5 FDDA system has been added to the new-release version of WRF.

1. INTRODUCTION

Fine particulate matter (PM2.5, referring to particles with aerodynamic diameters equal or less than 2.5 microns) has been implicated in a variety of health problems, including respiratory disease. With the recent decrease in the allowable 24-hour PM2.5 concentration to 35 micrograms per cubic meter, there is now an even greater need to be able to determine the sources primarily responsible for exceedance events when they occur, as well as to predict the potential impact of source emission changes. Modeling the behavior of fine particulate matter typically involves coupling between an inventory of emissions sources, chemical and air transport and dispersion models, and synoptic and mesoscale atmospheric models. (Synoptic atmospheric models are designed to represent features with characteristic horizontal scales greater than about 2000 km; mesoscale atmospheric models represent features with scales of approximately 2 – 2000 km.) The purpose of the meteorological models is to use physical predictive equations and assimilation of available meteorological data to capture the evolution of the local atmospheric state over sufficiently long periods for use by the other models.

During the winter season the part of interior Alaska consisting of Fairbanks and the surrounding Fairbanks North Star Borough often have extremely cold temperatures due to the strong longwave radiative cooling, the absence of moderating marine influences, and the generally weak winds. Although this region often has a clean, relatively pristine atmosphere, the periods of coldest temperatures are often accompanied by some of the strongest low-level temperature inversions that have been observed, with temperature increases up to 20°C as one ascends from the surface (Benson 1970). The inversions cap stable boundary layers (SBLs) that can be as shallow as tens of meters in clear nocturnal conditions (Sereze et al. 1992; Vickers and Mahrt 2004). Emissions from vehicular traffic, power plants, and home heating (mostly diesel and wood fuels) remain trapped within the SBL, leading to high concentrations of particulates and other pollutants. In the extremely cold conditions of interior Alaska an additional problem that arises is ice fog that can be triggered by combustion-generated water vapor at temperatures below approximately -25°C (Benson 1970; Girard and Blanchet 2001). The dispersal of pollutants is further hindered by the fact that winds and turbulence are quite weak in these conditions. The winds and turbulence that do exist in the SBL are strongly modulated by drainage flows, gravity waves, and other less understood phenomena (Hanna 1983; Mahrt 2009). Thus predicting the behavior of SBLs becomes a complex problem involving synoptic weather patterns, topography, turbulence, surface energy budgets, and precipitation.

The tool used for the meteorological modeling component of this project is the Weather Research and Forecasting (WRF) model (Skamarock et al. 2008), more specifically, the Advanced Research WRF dynamic core (WRF-ARW, henceforth simply called WRF). WRF contains separate modules to compute different physical processes such as surface energy budgets and soil interactions, turbulence, cloud microphysics, and atmospheric radiation. Since turbulent eddies in the SBL are typically much smaller than mesoscale model horizontal grid spacing (e.g., ten meters vs. a thousand or more meters), they cannot be modeled directly (e.g., Wyngaard 2004), but typically their effect is parameterized by a Planetary Boundary Layer (PBL) scheme that predicts turbulent kinetic energy (TKE). Within WRF the user has many options for selecting the different schemes for each type of physical process. There is

1

also a WRF Preprocessing System (WPS) that generates the initial and boundary conditions used by WRF, based on topographic datasets, land use information, and larger-scale atmospheric and oceanic models. New software associated with objective analysis and data assimilation will be discussed later.

The goal of this project was to select and perform two twenty-day simulations down to 1-km horizontal grid spacing for two episodes from the 2007-2008 winter season characterized by high PM2.5 exceedance events in the Fairbanks region. One episode was to be characterized by near total darkness, while the second was to contain partial sunlight. From a set of modeling experiments including a baseline physics configuration and a series of physics sensitivity tests, modified as appropriate to be suitable to the unique Alaskan atmospheric conditions, a best performing physics suite was to be selected and delivered to the EPA, along with source code and the model output. The project had two main components: (1) creating the best possible representation of the atmosphere through the use of a mesoscale model with continuous data assimilation, and (2) determining the best set of physics parameterizations by performing a series of sensitivity tests without the direct effects of data assimilation.   Both components are included in a multiscale, multigrid data assimilation procedure, which will be described in more detail below.

2. METHODOLOGY AND BASELINE EXPERIMENTAL DESIGN

2.1 Grid Configuration

The simulations presented in this report involve three one-way nested horizontal grids with horizontal grid spacing of 12 km, 4 km and 1.3 km, respectively (Table 1 and Fig. 1). Grid 1 covers the entirety of Alaska and extends from Siberia to the northwestern continental United States. Grid 2 closely coincides with the extent of the Alaskan landmass south of the Brooks range; it includes the Anchorage region and the Gulf of Alaska in the south. Grid 3, centered around Fairbanks and extending south to the Alaska Range and north past the White Mountains and other uplands just north of Fairbanks, includes all of the proposed non-attainment area within the Fairbanks North Star Borough (Fig. 2).   It can be seen in the figure that Fairbanks is located next to a semicircle of low mountains that are generally a few hundred meters above the city; this tends to restrict airflow near the city and further reduce the dispersion of pollutants in stable conditions.

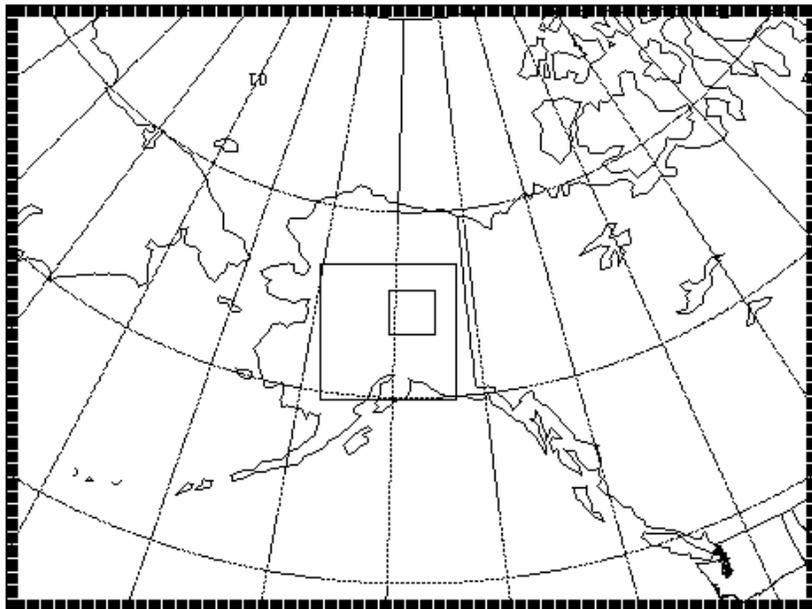| Grid	Dimensions	Horizontal Grid Spacing
1	401 x 301	12 km
2	202 x 202	4 km
3	202 x 202	1.3 km

Table 1: Specifications of model grids.

Fig. 1: Nested grid configuration of WRF, showing the 12-km Grid 1, the 4-km Grid 2, and the 1.3-km Grid 3 described in the text.
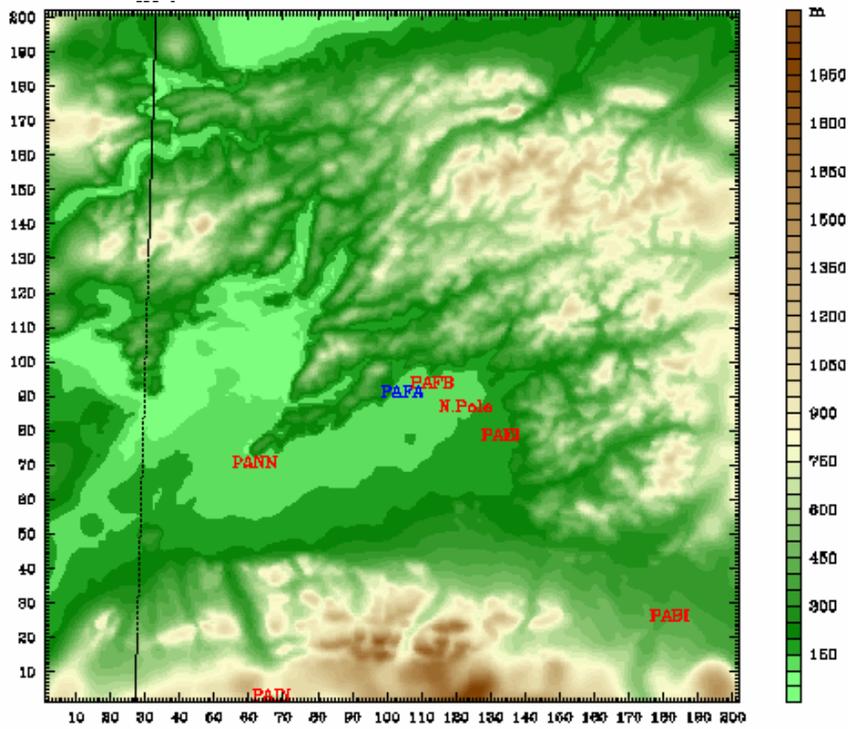

Fig. 2: Elevation on Grid 3 used in study. The location of the Fairbanks sounding is labeled in blue; other local METAR stations are shown in red.

The vertical grid spacing needed to be fine enough to resolve the structure of SBLs that can be only tens of meters deep, but not so fine that numerical instabilities arise in regions of steep topography (in particular the Alaska Range). After a series of initial tests a vertical grid configuration with 38 half layers (39 full levels) was defined, with a minimum vertical grid spacing of 4 m near the surface (see Fig. 3). Numerical stability was achieved through the use of time steps of 24 s, 8 s, and 4 s on the 12-km, 4-km and 1.3-km grids, respectively. These parameters are comparable to those used over central PA in the Seaman et al. (2008) SBL study. but with 4-m rather than 2-m vertical resolution near the surface, and slightly shorter timesteps.
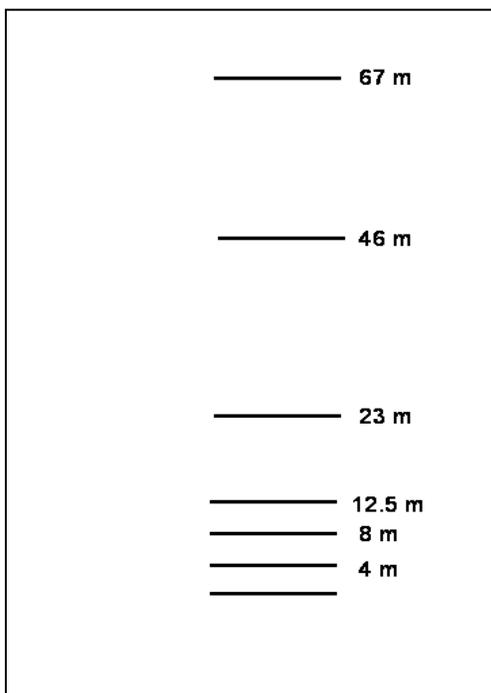
Fig. 3: Lowest few vertical full levels (i.e., locations where vertical velocity is calculated) in WRF model configuration, roughly to scale.

Two twenty-day episodes from the 2007-2008 winter season were selected for study. One episode was from 14 Dec 2007 to 03 Jan 2008, a time of year when there is little solar radiation in the Fairbanks area (approximately three hours of daylight per day near the solstice). During this episode the temperature rapidly decreased to near -40°C by 21 Dec, accompanied by rapid increases in PM2.5 concentrations, and then temperatures generally increased and PM2.5 decreased for the remainder of the episode (Fig. 4). The second episode was from 23 Jan 2008 to 14 Feb 2008, when solar insolation was more significant (between five and eight hours of sunlight per day), and provides an example of 'partial

sunlight' conditions. During this episode temperatures were initially relatively warm (near 0°C), decreased briefly to near -35°C by 27 Jan, rebounded slightly, and then decreased during the most extensive period of sub -35°C weather of the season. Consistent with the prolonged period of cold temperatures were recurring violations of the PM2.5 standard in the Fairbanks area.
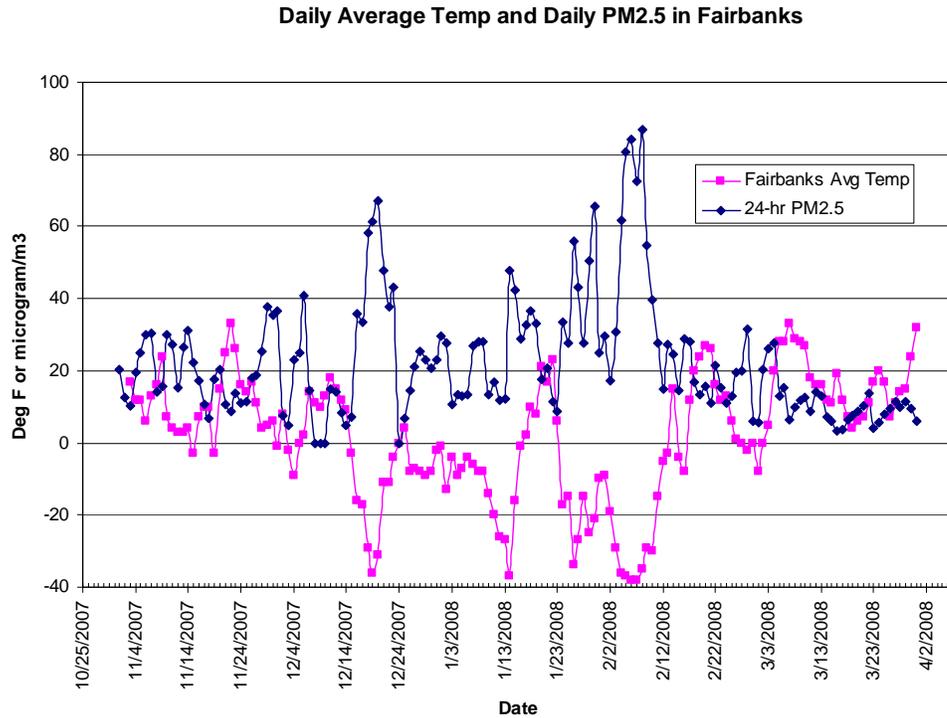
Daily Average Temp and Daily PM2.5 in Fairbanks

Fig. 4: Observations of daily average temperature and 24-hr PM2.5 concentrations taken in Fairbanks during 2007-2008 winter season. Courtesy Robert Dulla, Sierra Research, Inc.

In the initial period of a regional model simulation there is generally a period of several hours when the atmospheric state, whose initial conditions are usually provided by a global or coarser regional model, is still dynamically adjusting to the finer scale resolution and topography of the regional model. Therefore the model output from this initial 'spin-up' period is not completely reliable as an indicator of the true atmospheric state. However, if a regional model simulation is allowed to progress for too long without re-initialization (normally several days), it tends to drift away from the actual observed atmospheric state. Therefore, our method of obtaining realistic regional atmospheric analyses over an entire twenty-day episode was to divide each episode into four overlapping simulation segments. Each segment is around five days long with a twelve-hour overlap between each segment to avoid spin-up effects. (Specifically, the near total darkness episode was divided into successive segments of 6 days, 5.5 days,

5.5 days, and 4.5 days; the partial sunlight episode was divided into successive segments of 5 days, 5.5 days, 5.5 days, and 5.5 days).

Initial conditions and most of the Grid 1 lateral boundary conditions were obtained from the half-degree Global Forecast System (GFS) zero-hour analyses that were obtained from the NOAA National Operational Model Archive and Distribution System (NOMADS) website maintained by the National Climatic Data Center. The exceptions were some analysis times during the near total darkness episode when the half-degree GFS product was unavailable; in these instances the one-degree GFS analysis was used. All simulation segments for the near total darkness episode were selected such that all initial conditions could be obtained from half-degree global analyses.

The simulations were performed on one of two Linux clusters: one local cluster with 128 available processor cores, and the other cluster with 512 processor cores maintained by the Research Computing and Cyberinfrastructure High Performance Computing Group (RCC HPCG) at Penn State. Each 5.5 day simulation segment took 1-2 days to complete. The full 3D model output from each simulation was saved at a frequency of one hour for the 12-km Grid 1, and at a frequency of 12 minutes for the 4-km Grid 2 and 1.3-km Grid 3. For our configuration as shown in Table 1, the file size at each model output time is 500 MB for Grid 1 and 170 MB for each of Grids 2 and 3 (although this size can be approximately halved through file compression).

2.2 Four-Dimensional Data Assimilation (FDDA)

Even with the overlapping simulation segment strategy, it is difficult to ensure that the interior of a regional model simulation remains close to observations for simulations of more than a day or so. Therefore, dynamic analyses of historical cases are often performed, in which a Four-Dimensional Data Assimilation (FDDA) strategy is applied throughout the model integration. Relaxation terms based on the differences between actual observations and the corresponding model fields at the observation sites (also known as the 'innovations') are added to the model's predictive equations. In this way the model error is constrained based on available observations while the model still provides dynamic consistency and finer mesoscale structure not present in the observations. The version of FDDA used in these simulations is the multiscale, multigrid nudging FDDA strategy developed by Stauffer and Seaman (1994) for the MM5 mesoscale model. Nudging is also known as Newtonian relaxation, where the nudging relaxation terms are proportional to the innovation divided by a characteristic e-folding time inversely proportional to a nudging coefficient G. Nudging does not perform a direct insertion of observational information at a single point in space and time, but rather it applies the correction or innovation gradually in time and space based on the model terrain influences and prescribed / assumed weighting functions.   For example, when a well-mixed PBL is present, one would generally want the influence of surface observations to be extended throughout the PBL, because in these conditions there is high correlation between errors in atmospheric fields at the surface and those anywhere within the PBL.

The multiscale multigrid FDDA method uses a combination of two forms of nudging: analysis nudging and observation ('obs') nudging. Analysis nudging is performed in model grid space where an objective

analysis of observations (e.g., with a modified Cressman scheme (Benjamin and Seaman 1985)) is performed using the interpolated global analyses (e.g., from the GFS) as a background field. The resultant 'enhanced analysis' can then be used as the basis for analysis nudging. Analysis nudging is generally applied on coarser model domains where synoptic data can be used to produce a reasonable gridded analysis. Obs nudging is more attractive for finer-scale domains and asynoptic data. It is particularly effective where observational data density is sparse and corrections are applied only in the neighborhood of the observations, allowing the model to still add value in regions without any data by propagating observation information into the data-sparse regions and creating mesoscale structure not in the observations. In this case the nudging is performed in observation space, and the model field is interpolated to the observation site to compute the innovation that is then analyzed back to the model grid over some three-dimensional neighborhood in space, and over some time window. Quality control (QC) of observations is critically important for the success of both analysis nudging and observation nudging.

In the multiscale multigrid FDDA method applied in this study, 3D-analysis nudging, as well as surface analysis nudging using higher temporal frequency surface data within the PBL (e.g., Stauffer et al. 1991), are performed on the outermost 12-km domain. Obs nudging is applied on at least the 12-km and 4-km domains. (Obs nudging is not applied on the finest 1.3-km model nest for the physics sensitivity studies described further below.) The finer domains thus have the benefit of improved lateral boundary conditions from the coarsest 12-km domain using both types of nudging, as well as the obs nudging performed directly on the 4-km nested domain.

This project was one of the first applications of the multiscale FDDA strategy of Stauffer and Seaman (1994) in WRF. It is important to note that many of the WRF FDDA capabilities were not available and still under development via a contract from the Defense Threat Reduction Agency (DTRA) to Penn State at the time that this project was proposed. In fact, the WRF 3D / surface analysis nudging and obs nudging capabilities were still being developed during this contract period. The WRF end-to-end FDDA system is shown in Fig. 5 and described in more detail in Deng et al. (2009). This contract was able to take advantage of the fact that the WRF FDDA developers at Penn State were also working on this contract.

The new OBSGRID module in the WRF end-to-end FDDA system produces gridded objective analyses and observation files similar to those produced by Rawins / Little_r in the MM5 system. These files can be used for 3D/surface analysis nudging and obs nudging within WRF. OBSGRID takes as input raw WMO observations (both surface and upper air) and the output of WPS, which consists of atmospheric initial and boundary gridded data (e.g., GFS output) horizontally interpolated to the model grid to be used in WRF. The outputs of OBSGRID relevant to this study include 1) pressure-level and surface objective analyses of the WMO observations (passing internal QC checks) using the WPS GFS output as background fields; the resultant analyses are then vertically interpolated to the WRF terrain-following "sigma" layers to be used for 3D analysis nudging; 2) surface analysis nudging files that can be directly used by WRF; 3) observation nudging files usable by WRF, and 4) files of the WMO observations including those passing the QC tests for use in the statistical verification software.

As mentioned above, for the physics sensitivity part of this study, 3D analysis nudging, surface analysis nudging, and obs nudging are performed on the 12-km Grid 1; obs nudging is performed on the 4-km Grid 2; and no nudging is performed on the 1.3 km Grid 3. Thus Grid 3 has no direct FDDA tendencies and can be used to determine physics sensitivities, while still benefitting from improved lateral boundary conditions derived from the coarser grids that do have FDDA.
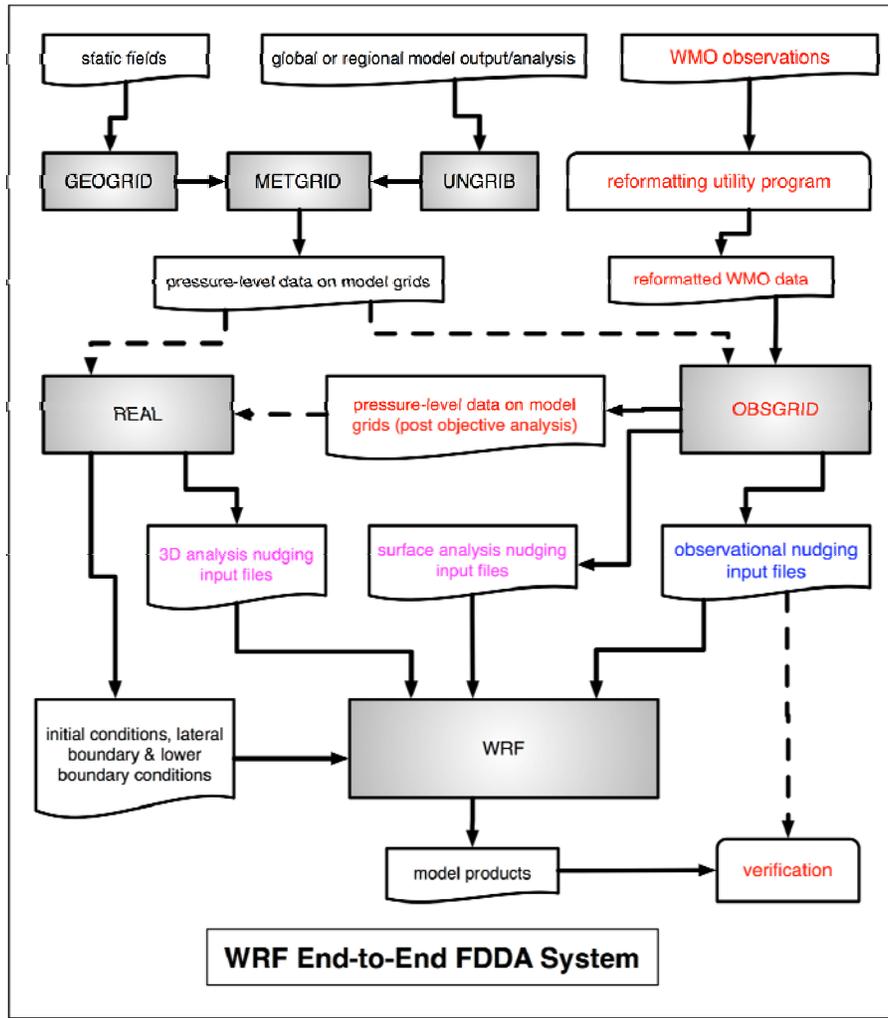

Fig. 5: Diagram of the WRF End-to-End FDDA system used for this study (from Deng et al. 2009). Items in magenta apply to analysis nudging; items in blue apply to obs nudging; items in red apply to both.

For the generation of the final dynamic analysis, obs nudging was performed on Grid 3, but with a reduced horizontal radius of influence (from 100 to 75 km), a reduced vertical pressure difference within the terrain-modified radius of influence function used for surface obs nudging (from 75 hPa to 37.5 hPa),

and obs nudging of surface data was performed on mass fields only (i.e., not winds). The values of FDDA-related WRF namelist parameters for these simulations can be found in Table 2.

	3D/Sfc Analysis Nudging			OBS Nudging		
Parameter	Grid 1 12-km	Grid 2 4-km	Grid 3 1.3-km	Grid 1 12-km	Grid 2 4-km	Grid 3 1.3-km
G (1/sec)	$3*10^{-4}$	**Not Used**	**Not Used**	$4*10^{-4}$	$4*10^{-4}$	$(4*10^{-4})$
Wind field	Yes	**Not Used**	**Not Used**	Yes	Yes	No
Mass field	Yes	**Not Used**	**Not Used**	Yes	Yes	(Yes)
RINXY (km)	N/A	N/A	N/A	100	100	(75)
TWINDO (hours)	N/A	N/A	N/A	2	2	(2 – but see Section 5)
Time Frequency of Data (hours)	6 / 3 (Sfc)	**Not Used**	**Not Used**	1	1	(1)

Table 2: List of WRF FDDA namelist parameter values used in this study. Analysis nudging parameters apply to both surface and 3D versions unless otherwise specified. Values in parentheses for Grid 3 do not apply to the physics sensitivity studies, which have no FDDA on Grid 3, but do apply to the final dynamic analysis performed in this study.

2.3 Baseline Physics Suite

Two of the most important controls on the evolution of SBLs in mesoscale models are the PBL scheme and the Land Surface Model (LSM). The former is critical for determining the effects of vertical mixing both within and outside of the PBL, and thus helps regulate how rapidly pollutants can disperse. The LSM helps to determine the details of the surface energy balance and thus the thermal tendency and stability of air near the surface. In addition to these, other physical processes that are important in these conditions are the atmospheric radiation scheme (because of the impact on the thermal cooling and temperature structure of the lower atmosphere) and the microphysics scheme (because of the interactions between radiation, latent heat, and quantities of water vapor and condensate, as well as the value of predicting such features as ice fog).

The baseline physics suite used for these simulations was originally derived from that of Seaman et al. (2008) for central Pennsylvania, but with some modifications. To determine the longwave component of radiation, the RRTM scheme of Mlawer et al. (1997) was used, whereas the Dudhia (1989) scheme was used to determine the shortwave component. The PBL scheme used was a version of the Level 2.5 Mellor Yamada scheme as modified by Janjić (2002); henceforth this will be referred to as the Mellor-Yamada-Janjić (MYJ) scheme. A Level 2.5 scheme explicitly predicts the evolution of turbulent kinetic energy (TKE) at each grid point, and uses the predicted TKE to compute the magnitude and vertical extent of mixing. The MYJ scheme used is that available in version 3.1 of WRF; however, based on subsequent work from the central Pennsylvania study, the threshold of minimum TKE within the MYJ scheme was reduced to 0.01 $m^2 s^{-2}$, due to the extremely weak winds and turbulence expected in these stable conditions.

The LSM for the baseline was originally the 5-layer thermal diffusion model used in Seaman et al. (2008). However, we performed a series of preliminary tests with the Alaska grid configuration using the Noah LSM, originating from NCEP, Oregon State University and AFWA (Chen and Dudhia 2001). This was done because the Noah LSM includes a number of features that are potentially important in the central Alaska environment, including time-dependent snow cover, time-dependent snow density, and snow-dependent emissivities and ground conduction. Some properties of the Noah LSM that had just been incorporated into standard WRF (e.g., a more rigorous treatment of latent heat release in the presence of ice) were based on the 'Polar-WRF' and 'Polar-MM5' versions of Noah used for high latitude simulations (Bromwich et al. 2001; Hines and Bromwich 2008). A number of other features of the polar-modified Noah were not in the standard WRF at the time, but not directly relevant to central Alaska (e.g., modification of sea ice properties). Preliminary tests in the relatively mild conditions immediately prior to the partial sunlight episode revealed that the use of the Noah LSM initialized directly from the soil levels of the half-degree GFS resulted in smaller surface temperature biases. Thus, based on our preliminary favorable results, we used the version of Noah in WRF v3.1 as the LSM for the baseline simulation.

The microphysics model selected for the baseline was the Morrison et al. (2005) scheme, also new to WRF v3.1. This scheme was developed specifically for high-latitude cold temperature microphysics, and includes the prediction of two moments (mixing ratio and number concentration) for rain, snow, graupel, and cloud ice, in addition to single moment prediction of cloud water. We thus felt it was worth using this scheme in the baseline even though file sizes and computational costs were significantly increased (by 50% in time) from the simple ice scheme used previously.

3. Initial Baseline Testing and FDDA Modifications

Initial testing of the baseline WRF configuration for the two episodes began in January 2009. The purpose of the 'pre-baseline' testing was to confirm that the proposed WRF grid configuration would remain numerically stable and physically realistic for simulation segments of several days, to determine the resource and timing requirements of the simulations, and to confirm that the WRF FDDA features were working as expected. Furthermore, a number of key WRF system features to be used in this study

were still under development at the beginning of 2009; in particular, surface analysis nudging, OBSGRID, and the official WRF v3.1 release itself, which included the QNSE PBL scheme and a modified version of the Noah LSM. Thus all of these new features had to be tested and evaluated when they became available.

At the beginning many of these tests were performed on the first segment of the partial sunlight episode (23 – 28 Jan 2008). Not only was this a convenient place to begin, but it began as a time of relatively warm temperatures in central Alaska, allowing the model configurations to be evaluated in relatively mild conditions before being used in the extreme cold conditions of the high exceedance episodes. Nonetheless, a brief period of colder temperatures occurred toward the end of the 23-28 Jan 2008 period, so some evaluation of model performance in different temperature regimes could be determined.

A preliminary assessment of the skill of the FDDA components of the WRF end-to-end system for the baseline simulation of the 23-28 Jan 2008 period, made in April 2009, is shown in Table 3 for the 12-km (Grid 1) and 4-km (Grid 2) domains. Raw WMO observations from both surface METAR and rawinsonde stations were given QC codes within OBSGRID, and only those observations of sufficient quality to be used in the objective analysis were retained for verification. The table compares a simulation without FDDA, a simulation using only analysis nudging on Grid 1; a simulation using only obs nudging on Grids 1 and 2; and a simulation combining the analysis nudging and obs nudging features, corresponding to the proposed multiscale multigrid FDDA procedure. Furthermore, since the surface analysis nudging feature of WRF had only just become available from Penn State, two versions of each simulation including analysis nudging were performed: one with and one without surface analysis nudging.

The table confirms that, for virtually every grid, observation station type, and variable, the best root-mean-square error (RMSE) scores occur for multiscale multigrid FDDA, and the worse RMSE scores occur for the simulation without any FDDA. However, a more careful analysis of the table revealed a few puzzling results. While surface analysis nudging led to expected improvements in temperature on Grid 1 (vs. analysis nudging without surface analysis nudging) when verified against surface METAR stations, the RMSE scores of METAR winds and relative humidity actually became slightly worse. Furthermore, when the verification was performed against rawinsondes on Grid 1, surface analysis nudging made temperature RMSEs considerably worse, and wind RMSEs far worse, than the corresponding runs without surface analysis nudging.

For Grid 2 verified against rawinsonde data, we see the expected result that a simulation with only obs nudging improves the RMSE scores more than either version of the analysis nudging only simulation. (Since analysis nudging is always applied to Grid 1 only, the analysis-nudging-only simulations have only indirect FDDA improvements on Grid 2, through the lateral boundary conditions from Grid 1; the obs nudging simulations do have direct FDDA on Grid 2.) However, when surface METARs are used for Grid 2 verification, we have the puzzling result that obs nudging only is outperformed by analysis nudging only (except for temperature).

Verification Domain	Verification Field and Station Type	Simulation FDDA Method (O – Obs Nudging; 3DA – 3D Analysis Nudging; SA – Surface Analysis Nudging; No – No Nudging)			
		Grid 1: No Grid 2: No	Grid 1: O Grid 2: O	Grid 1: 3DA / 3DA + SA Grid 2: No / No	Grid 1: 3DA + O / 3DA + SA + O Grid 2: O / O
Grid 1 (12 km)					
	Surface U-Component	3.2	2.6	2.3 / 2.4	**2.1** / 2.2
	Surface V-Component	3.2	2.7	2.1 / 2.3	**2.0** / 2.1
	Surface Temperature	5.6	2.9	2.9 / 2.4	2.5 / **2.1**
	Surface Rel. Humidity	21.0	18.7	17.7 / 18.2	**17.0** / 17.5
	Sounding U-Component	4.6	2.2	1.5 / 3.3	**1.1** / 2.0
	Sounding V-Component	4.2	2.3	1.5 / 2.9	**1.1** / 1.9
	Sounding Temperature	3.5	1.4	1.4 / 2.0	**1.0** / 1.3
	Sounding Rel. Humidity	21.2	10.2	11.2 / 16.0	**8.3** / 10.5
Grid 2 (4 km)	Surface U-Component	3.8	3.3	**2.2** / 2.3	2.5 / 2.7
	Surface V-Component	**2.5**	3.1	2.7 / 2.8	2.9 / **2.5**
	Surface Temperature	5.0	2.5	3.1 / 3.0	1.9 / **1.8**
	Surface Rel. Humidity	23.8	22.0	20.7 / 20.7	19.6 / **19.3**
	Sounding U-Component	4.5	2.2	2.6 / 2.8	**1.7** / 1.8
	Sounding V-Component	4.5	3.2	3.4 / 3.8	**2.8** / 3.4
	Sounding Temperature	3.1	1.3	2.2 / 2.2	**0.9** / 1.4
	Sounding Rel. Humidity	27.0	14.1	21.7 / 24.5	**12.5** / 13.1

Table 3: Root-mean-square error (RMSE) values of u-component wind (m s^{-1}), v-component wind (m s^{-1}), temperature (°C) and relative humidity (%) as verified within Grids 1 and 2 during test FDDA simulation of 23-28 Jan 2008 for various FDDA combinations. Verification was performed against METAR stations for the surface and rawinsonde stations for the sounding data. The best value for each row is in bold.

Investigations into the cause of these puzzling results led to the realization that a number of the components of the WRF end-to-end FDDA system probably needed to be modified to adapt the system to the special conditions of the Alaska configuration. First, in most mesoscale model simulations it can be assumed that surface wind observations, normally made at a height of 10 m above ground level (AGL), and surface temperature and moisture observations, normally made at 2 m AGL, are located within the lowest model layer. In fact, normally the problem is that the midpoint of the lowest model layer (or first half-layer height above the surface) is often tens of meters in height and still well above the height of the surface observations. A proper interpolation of model values to the height of the surface observations usually requires using similarity theory or some similar procedure. For the Alaska configuration, however, a 10-m wind would actually be located within the *third* model layer from the surface, while 2-m temperature essentially corresponds to the height of the lowest model half layer (midway between the surface and the lowest model full level). There are at least two consequences of this. The first is that, for the default procedure of verifying surface wind observations with model output from the lowest model half layer, observed 10-m winds are actually being compared to modeled 2-m winds whereas they should be verified against the modeled 10-m winds of the third model half layer. The second consequence is that the surface wind innovations used in the WRF FDDA code are by default based on the difference between 10-m observed winds and 2-m modeled winds in this case, which is wrong and may introduce erroneous biases into the FDDA simulation.

An additional issue was revealed by examining fields of PBL height produced by the PBL turbulence parameterization in various test simulations. Though, as expected, PBL heights are very low over many large areas within the model domains, especially during the colder periods, some patches of unexpectedly high PBL height values can be seen at times (Fig. 6). PBL heights of 1500 m or greater are more typical of convective boundary layers than of the nocturnal SBL conditions found in interior Alaska. Model soundings taken in the proximity of these patches (Fig. 7) confirm that the atmosphere is certainly rather stable and not well mixed in potential temperature (although some layers above show potential temperatures close to a saturated adiabat). The high PBL height zones appear to be associated with regions of elevated shear-generated TKE and cloudiness, since it is the TKE profile in the MYJ scheme that determines the PBL height. The issue is that the default WRF surface analysis and obs nudging schemes spread the influence of surface innovations throughout the depth of the PBL, but in these stable conditions this may overestimate the vertical error correlation length scale for surface innovations. This helped explain why the use of surface analysis nudging on Grid 1 made the rawinsonde-verified statistics worse.
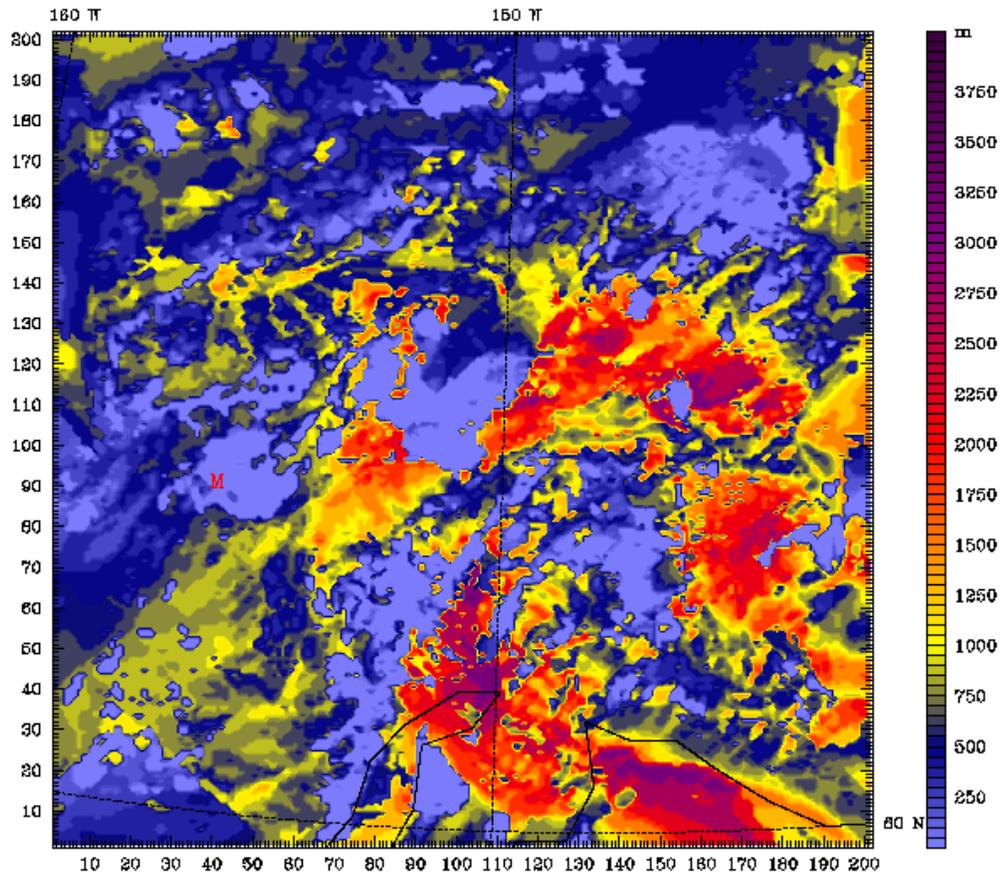
Fig. 6: WRF-predicted PBL height at 1200 UTC 25 Jan 2008 (60-hour simulation time) within the 4-km Grid 2. Simulation does not include FDDA.
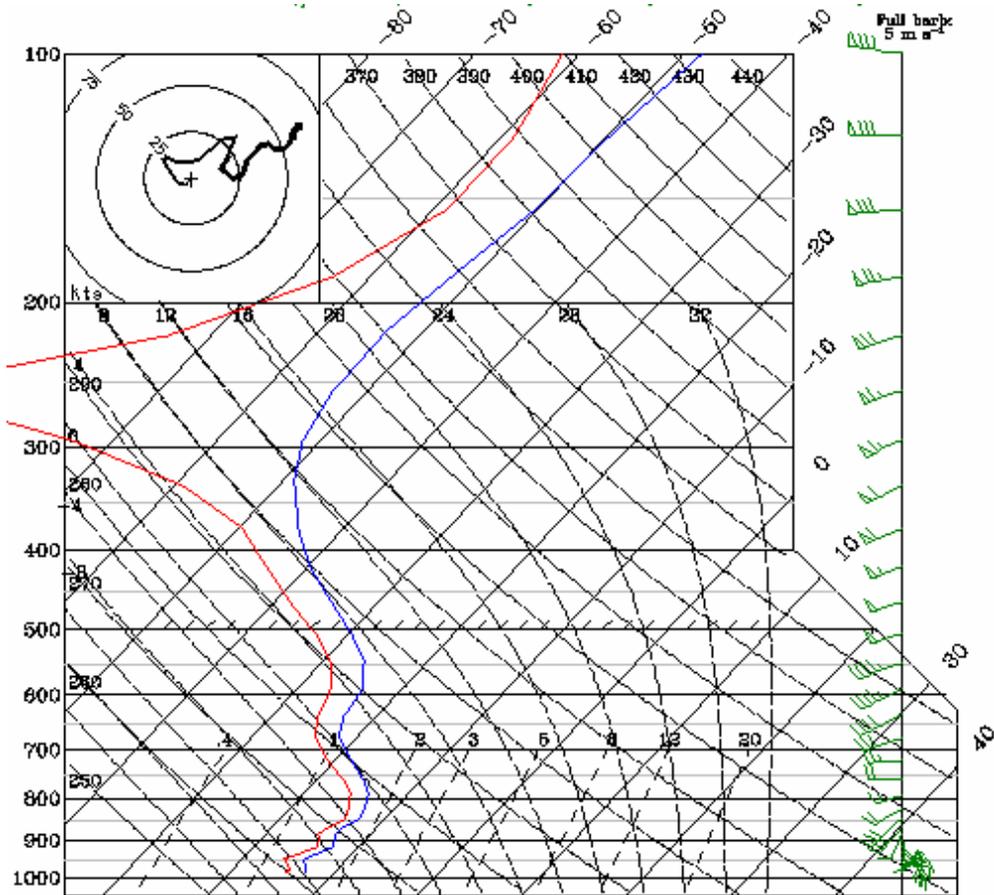
Fig. 7: WRF-predicted model sounding at
Fairbanks for same time and simulation as Fig. 6.

Based on a series of similar tests, the following modifications were made to the WRF FDDA schemes for use in the baseline Alaska simulations. 1) The verification software was rewritten so that surface wind observations are verified against the third model half-layer from the ground, while surface moisture and temperature observations are verified against the lowest model half-layer. 2) A portion of the verification software that uses an assumed lapse rate to adjust model temperatures based on the difference between modeled and actual elevation was disabled, because this can lead to large errors in very stable conditions. 3) The surface analysis nudging and obs nudging codes were modified so that surface innovations for wind are computed and applied directly at the third model level. 4) Because surface wind observations directly relate to the third model layer and surface temperature and moisture observations directly relate to the lowest model layer, the similarity-based adjustments normally performed on model output for surface innovation computation was also disabled. 5) Hardwired vertical weighting functions for surface innovations were implemented into the surface analysis nudging and obs nudging codes, replacing the default functions that extend surface corrections to the model-predicted PBL height. Trial and error established that the functions shown in Fig. 8 for surface obs

nudging and analysis nudging extend the surface innovations in the vertical enough to improve surface statistics but without degrading rawinsonde-verified RMSE scores; furthermore, the vertical extent of
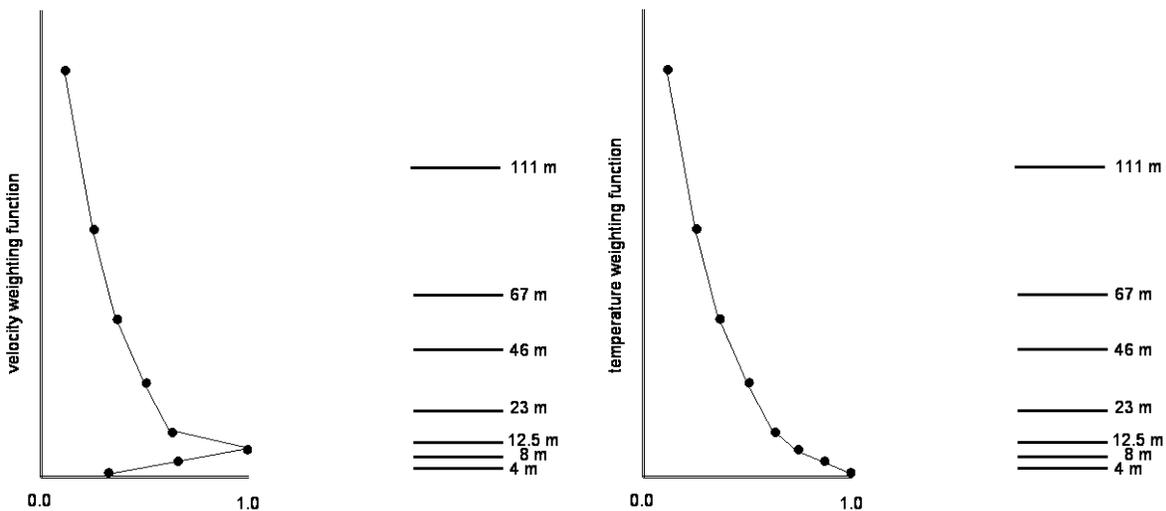

Fig. 8: Vertical weighting functions at model half-layers for wind components (left) and temperature and moisture (right), as used by modified surface analysis nudging and obs nudging FDDA procedures for Alaska simulations. Heights of model full layers are shown to the right, roughly to scale.

these functions (about 150 m) is a reasonable order of magnitude estimate for the maximum depth of nocturnal radiatively-driven SBLs.

Results from this phase of the project were presented at an oral presentation at the 13th AMS Conference on Mesoscale Processes in Salt Lake City, UT, from 17-20 Aug 2009. (Gaudet et al. 2009).

4. PHYSICS SENSITIVITY EXPERIMENTS

4.1 Experimental Design

Three modifications of the baseline physical parameterizations were evaluated in the physics sensitivity component of this project. The first involved modifying the atmospheric radiation schemes so that both the longwave and shortwave components used the new RRTMG radiation package, which uses the RRTM methodology but in a more efficient form adaptable to global climate models. This particular radiation package first became available in WRF v3.1. Though the RRTM and RRTMG longwave radiation schemes should produce very similar clear-sky fluxes, when mult-layered condensate is present the RRTMG makes use of the Monte Carlo Independent Column Approximation (McICA) to take into account 3D scattering effects.

The second involved changing the PBL parameterization from MYJ to the Quasi-Normal Scale Elimination (QNSE) scheme (Sukoriansky et al. 2005; Galperin et al. 2007). The theory behind the scheme is quite

advanced, but it is specifically designed for stable conditions, and allows both turbulent mixing and gravity wave motions to be represented in a unified framework. Dr. Boris Galperin was invited to Penn State University to give a seminar on the theory of the QNSE scheme in October 2008 before it was officially made public in WRF v3.1. The implementation of the QNSE scheme in WRF is actually similar to that of the MYJ, but with the values of vertical mixing parameters derived from the theory as a function of Richardson number (i.e., essentially the ratio of atmospheric stability to the square of the wind shear).

The third modification involved changing the LSM model from Noah to the Rapid Update Cycle (RUC) LSM. Among the features of the RUC LSM that suggest its use for this study is the presence of a snow model that potentially can have multiple layers depending on the snow depth (Smirnova et al. 2000). Other users have reported favorable results from using the RUC LSM in simulations of the Arctic (Mölders and Kramm 2010). The RUC LSM can also be initialized using soil information from the half-degree GFS after minor modification of the WRF source code. By default WRF can use either 6 or 9 soil levels, but we chose 6 because it is closer to the 4 levels of Noah and because it is the typical number of soil levels used in the RUC (e.g, Hines and Bromwich 2008).

4.2 Model Initialization and Setup

The objective analyses used for model initialization and analysis nudging were performed using the multi-quadric method within the OBSGRID software designed for WRF.   The background analysis files were derived from the half-degree GFS and topographic and land use dataset through the WPS. The background fields also served as the basis for performing QC on the WMO rawinsonde and surface METAR data used for verification and obs nudging, through 'buddy-check' (excluding obs too different from their neighbors) and 'err-max' (excluding obs too different from the background) procedures. A consequence of the current QC methodology is that all observations were located at the surface or at the standard pressure levels of the GFS model.

For the baseline and sensitivity experiments the model setup was the same except for the choice of physics options. Both the near total darkness and the partial sunlight episodes were simulated in their entirety using the four overlapping simulation segments referred to above. The FDDA procedure (using the modified vertical weighting functions) was defined to use surface and 3D analysis nudging on the 12-km Grid 1, obs nudging on both the 12-km Grid 1 and the 4-km Grid 2, and no FDDA on the 1.3-km Grid 3. Physics sensitivities on Grid 3 would thus be given greater weight than sensitivities on the other grids (which would not be expected to be as large due to the influence of FDDA). (However, we left open the possibility of performing a final dynamic-analysis simulation with obs nudging also performed on Grid 3 once a best-choice physics suite was selected; this final simulation would then have our best available model analysis of the atmospheric state during the episodes, and would be appropriate for use in atmospheric chemistry or transport and dispersion models. These results have been added to the report in Section 5 below.)

17

For each sensitivity experiment verification was performed using model output every 3 hours (excluding the initial time). For the periods of overlapping simulation segments, the model output from the segment at the larger forecast time was used, so all of the verification model output was at least 12 hours after a model initialization (except of course for the first 12 hours of an episode). All three grids were verified against only those stations located within the boundaries of Grid 3, to ensure that statistical differences between grids are not due to the different set of stations available on each domain. As previously discussed, verification of surface METAR data is performed directly with the third model level from the surface for winds, and the lowest model level for temperature and moisture.

The first physics sensitivity test involved changing the radiation to the RRTMG scheme for both longwave and shortwave components. We all agreed that if this produced favorable results we could simply retain the RRTMG radiation scheme rather than the Dudhia shortwave / RRTM longwave radiation suite of the baseline simulation in future sensitivity experiments. An initial three-day test (23-26 Jan) was performed without FDDA on any grid so as to maximize physics sensitivity. It was indeed found that the surface METAR temperature RMSE scores were consistently improved by the use of RRTMG (Fig. 9), although winds and relative humidity were little affected (not shown). The improvement seemed to be related to reduced downward longwave fluxes beneath patches of ice condensate. Thus, the decision was made that all future physics sensitivity tests, this time with FDDA on Grids 1 and 2 as described above, would make use of the RRTMG scheme.

The combinations of physics parameterizations used in the physics sensitivity tests are summarized in Table 4. To facilitate the comparison of different physics sensitivity experiments, the baseline simulation, with the combination of MYJ PBL scheme, Noah LSM, and Dudhia shortwave / RRTM longwave radiation, will henceforth be denoted as experiment MND. Another experiment, with MYJ PBL / Noah LSM / RRTMG radiation, will be noted as MNR, and another with QNSE PBL / Noah LSM / RRTMG radiation will be denoted as QNR. Finally, the experiment with MYJ PBL / RUC LSM / RRTMG radiation will be denoted as MRR.

Experiment Name	Planetary Boundary Layer (PBL)	Land Surface Model (LSM)	Radiation
MND (Baseline)	Mellor-Yamada-Janjić (MYJ)	Noah	Dudhia Shortwave / RRTM Longwave
MNR	Mellor-Yamada-Janjić (MYJ)	Noah	RRTMG Shortwave / RRTMG Longwave
QNR	Quasi-Normal Scale Elimination (QNSE)	Noah	RRTMG Shortwave / RRTMG Longwave
MRR	Mellor-Yamada-Janjić (MYJ)	Rapid Update Cycle (RUC)	RRTMG Shortwave / RRTMG Longwave

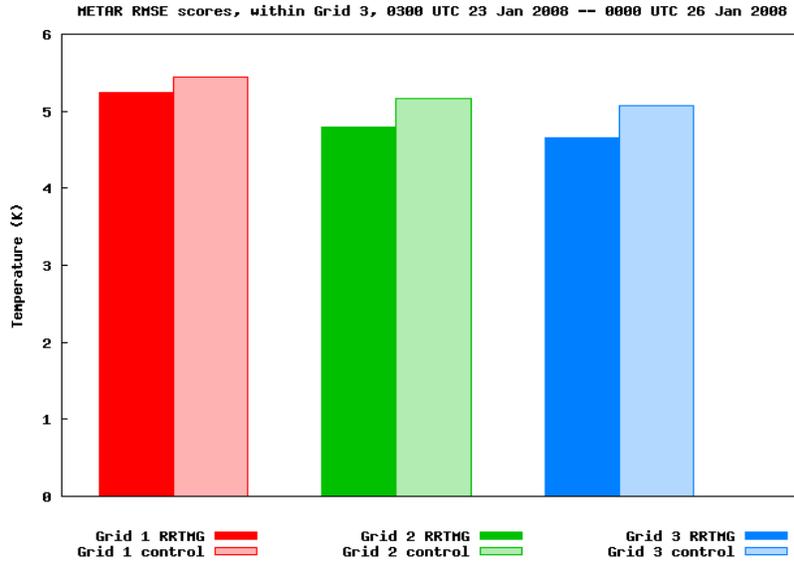*Table 4: Names and physical parameterizations used for physics sensitivity studies.*

Fig. 9: Surface METAR RMSE scores for temperature compiled for those stations located within Grid 3 for simulations from 00 UTC 23 Jan 2008 – 00 UTC 26 Jan 2008. Verification statistics are computed every three hours during the period. 'Control' denotes baseline physics configuration; 'RRTMG' denotes baseline physics configuration but with the RRTMG longwave and shortwave radiation schemes. All simulations shown were performed without FDDA.

4.3 Results of Physics Sensitivity Experiments

Figures 10 and 11 present the temperature RMSE and bias scores, respectively, for Grid 3 surface METAR stations for both the partial sunlight and near total darkness episodes. First, it can be seen that the RMSE score increases from Grid 1 to Grid 2 to Grid 3, which can be explained by the fact that less FDDA forcing is being applied from Grid 1 (both analysis and obs nudging) to Grid 2 (obs nudging) to Grid 3 (no nudging). These RMSE scores are large compared to typically reported surface meteorological values (e.g., Seaman and Michelson 2000), but of course the large temperature range through the period (about 40°C for both episodes) and extreme conditions make these challenging forecasts for a numerical model. Second, we see the previously discussed result that switching the radiation to RRTMG (compare MND and MNR) leads to improved temperature RMSE scores and lower positive temperature biases; the improvement is most noticeable on the no-FDDA Grid 3. The fact that the RMSE improvement through the use of the RRTMG is greater for the near total darkness episode than for the

partial sunlight episode was not unexpected, because previous examination of the partial sunlight episode revealed that the reduced positive temperature bias with RRTMG was due to the longwave component while the shortwave component of RRTMG partially counteracted this effect (not shown).
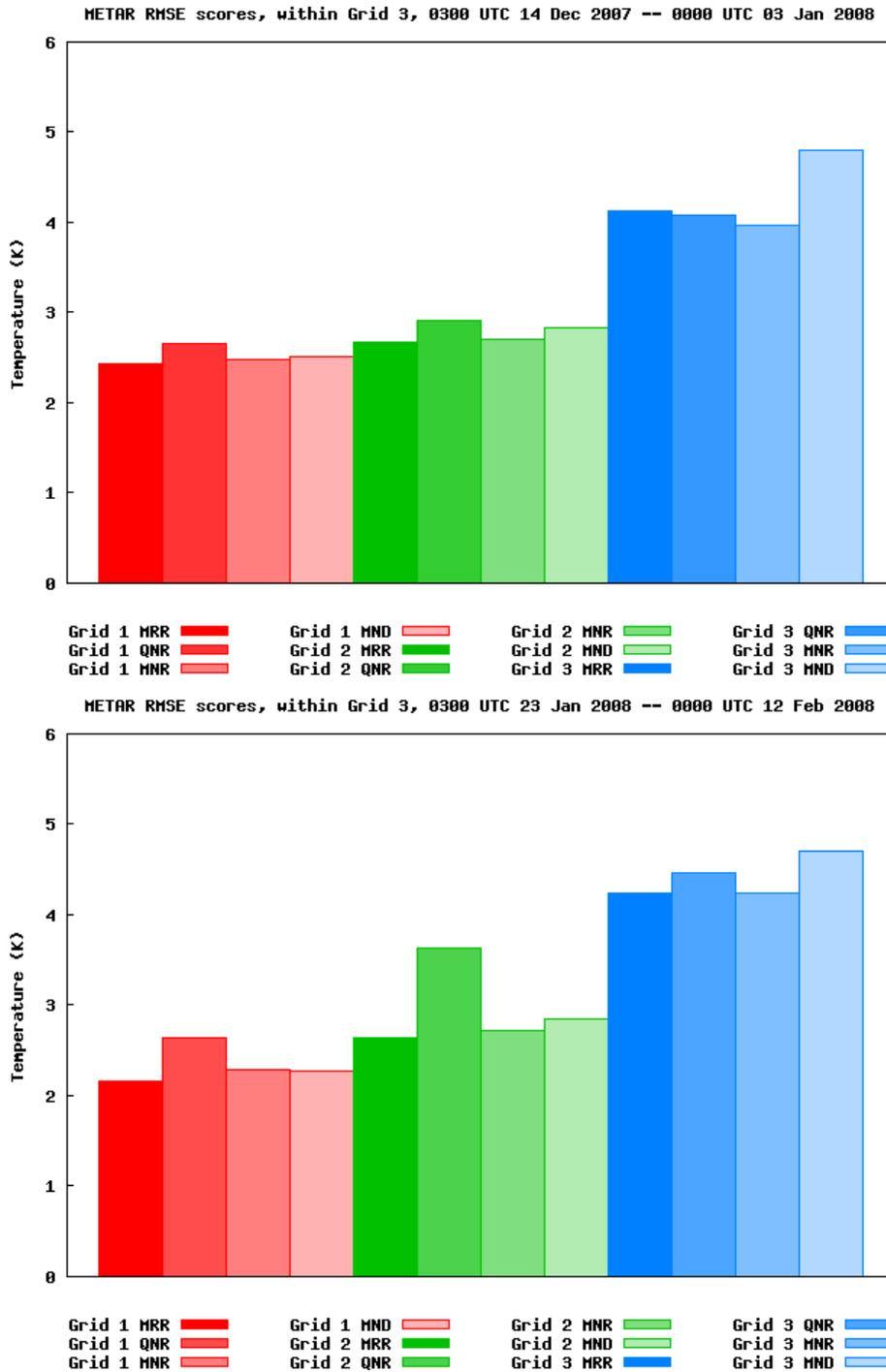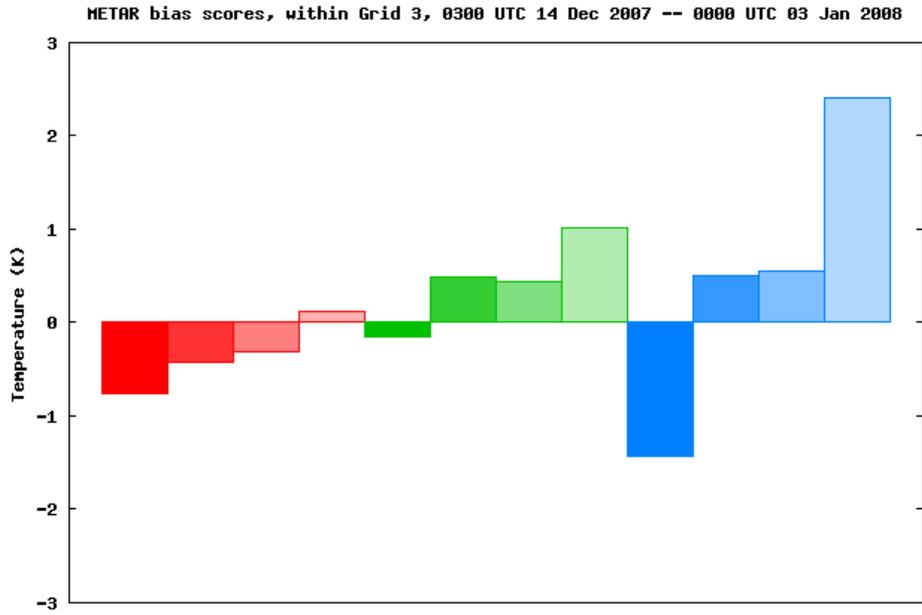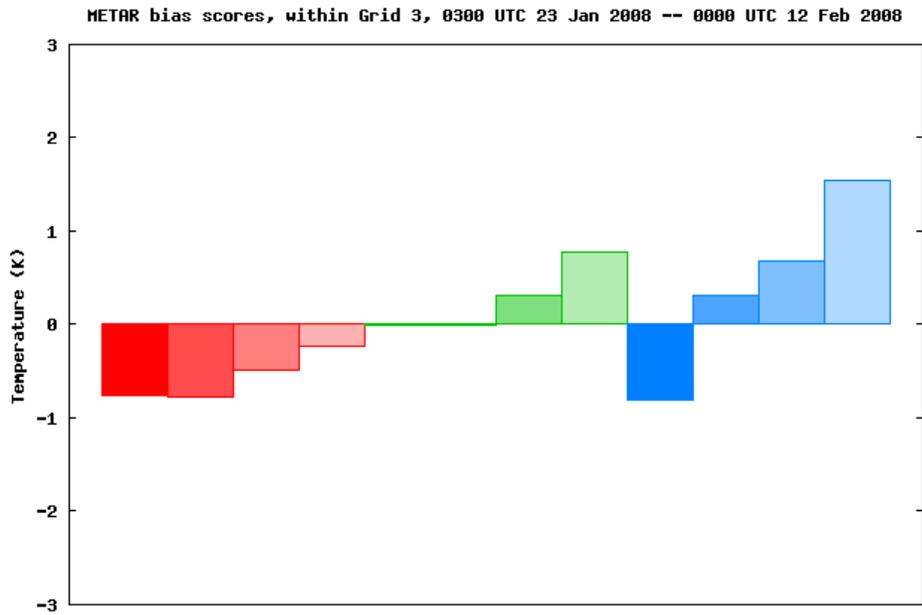

Fig. 10: Surface METAR RMSE scores for temperature for entire near total darkness episode (top) and partial sunlight episode (bottom). Labels for degree of shading refer to experiment names in Table 3. Verification statistics were computed every 3 hours during each episode as described in text.

Fig. 11: Same as Fig. 10, but for surface METAR bias scores for temperature.

Switching from the MYJ to the QNSE PBL scheme (compare MNR to QNR) further reduced and improved the magnitude of positive METAR temperature bias (for Grid 3 and the Grid 2 partial sunlight episode). However, the temperature RMSE scores for QNR are consistently greater than those for MNR; so this improved bias is not reflected in more skillful forecasts. The results of the QNSE PBL scheme are encouraging and should be analyzed in greater depth in a future project. We decided that the sensitivity test introducing the RUC LSM should use the MYJ scheme due to our greater experience with MYJ in WRF.

The effect of switching from Noah to RUC (compare MNR to MRR) produces the coldest surface temperatures of any of the experiments. While this leads to the best magnitude METAR temperature biases for Grid 2, the MRR Grid 3 temperature bias is considerably more negative, especially for the near total darkness episode. The MRR temperature RMSE scores for the METARs are the best, or tied for the best, of the four physics experiments for Grid 2 and the Grid 3 partial sunlight episode, but slightly worse than MNR and QNR for the Grid 3 near total darkness episode.

In terms of surface METAR wind speed RMSE and bias errors (Figs. 12 and 13) we see that there is less variability among the different physics schemes. For virtually all variables, grids, and episodes, however, the scores for experiment MNR are slightly better than the others. The wind speed RMSE scores tend to be slightly worse on Grid 3 without FDDA than on Grid 2 with obs nudging, but better than those on Grid 1 that uses analysis nudging but with a much coarser horizontal resolution.
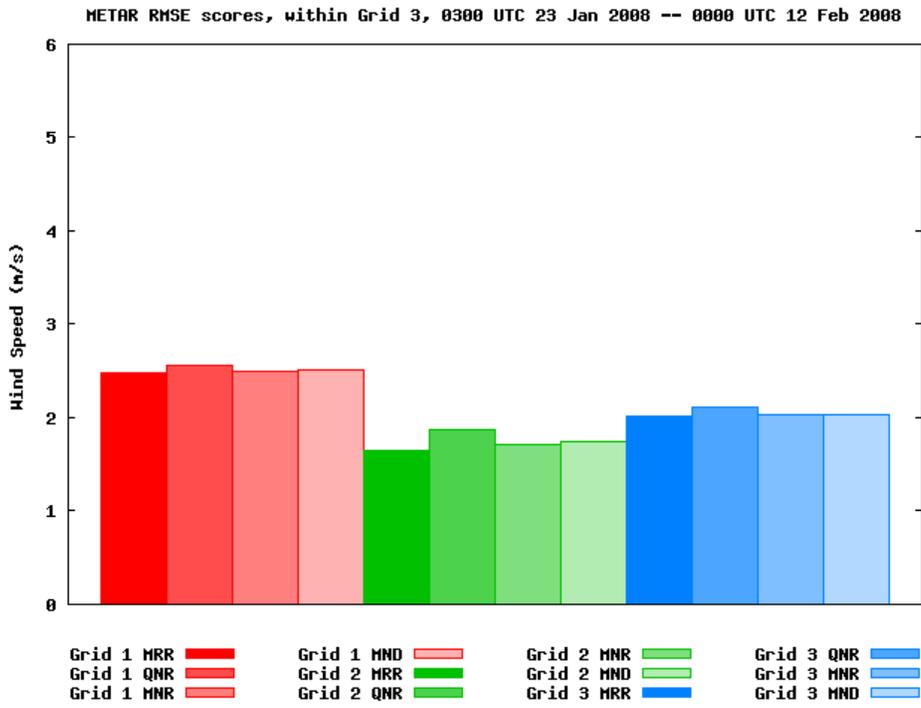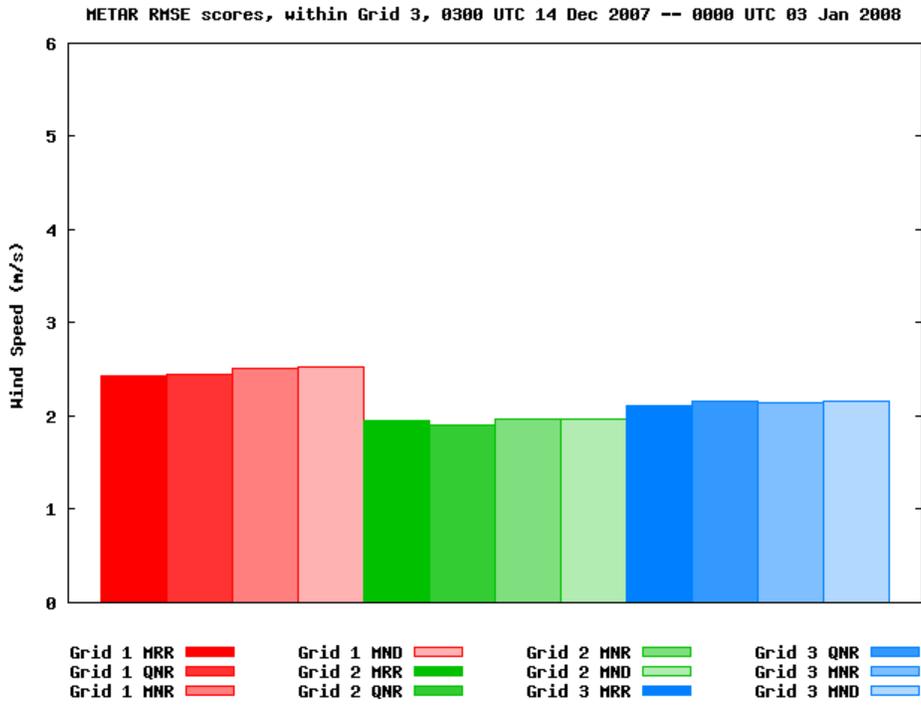
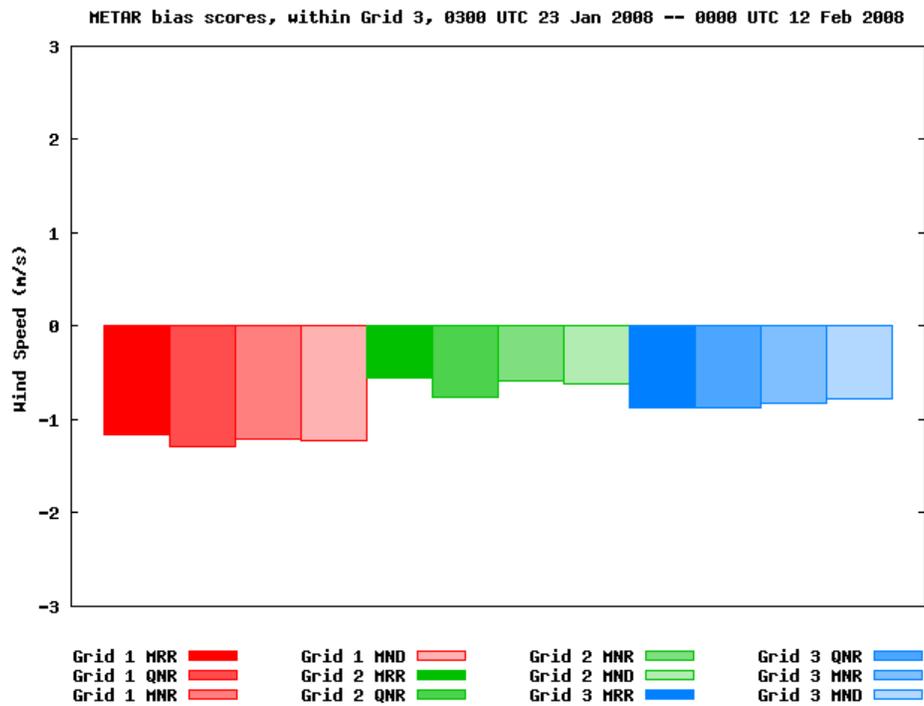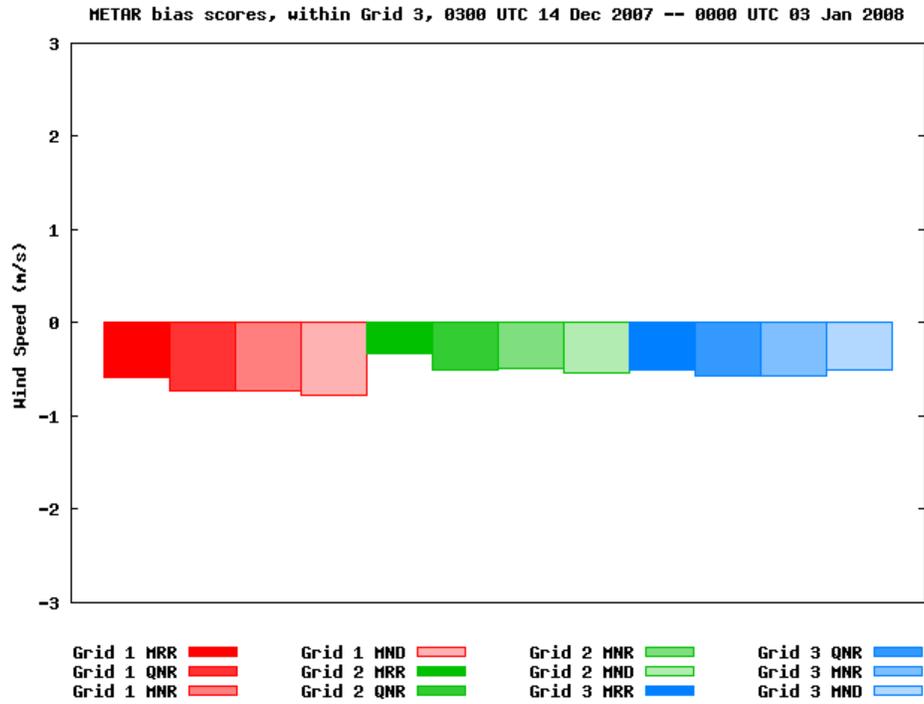Fig. 12: Same as Fig. 10, but for surface METAR RMSE scores for wind speed.

Fig. 13: Same as Fig. 10, but for surface METAR bias scores for wind speed.

In order to learn more about the nature of these biases, statistics for each episode can be compiled for each of the four 5-day (or 5.5-day or 6-day) simulation segments. One instructive comparison is between 14-20 Dec 2007 and 20-25 Dec 2007 (Fig. 14). The temperature difference between different physics configurations remains quite consistent between the two periods, but in the 14-20 Dec period the model temperature bias tends to be more negative than for the 20-25 Dec period. It can be shown that the large negative temperature biases of MRR are predominantly from the 14-20 Dec period. It should be noted that the highest exceedances / lowest temperatures for the near total darkness episode occur around 21 Dec. In general, Grid 3 tends to magnify the temperature biases of Grid 2, except for the MRR model for 20-25 Dec, where the Grid 3 temperature bias is reduced to almost zero.

A time series of the averaged observed, MNR, and MRR temperatures at the locations of the Grid 3 METARs is shown in Fig. 15. Note that the strongest MRR negative temperature biases in each episode tend to occur during times when the temperature is decreasing toward the coldest temperatures near 21 Dec in the near total darkness episode and about 05-09 Feb 2008 for the partial sunlight episode. At these times the MNR temperature bias also tends to be negative, but not by as much. However, immediately after the coldest temperatures are reached, the model biases become positive, and then the MRR configuration is preferred because temperature biases are smaller in magnitude. In particular, during the cold 5-9 Feb period MRR lacks the strong positive spikes in temperature bias that occur in the MNR simulation during the afternoons.

Verifying model features above the surface was made difficult by the scarcity of such observations in the region; the only rawinsonde sounding stations within Grid 2 are at Anchorage, McGrath, and Fairbanks, and of these only Fairbanks is located within Grid 3. No other reliable set of above-surface observations within Grid 3 was available for the episodes. A time-averaged composite of the vertical temperature structure of the Fairbanks sounding, compared to that from the different model physics configurations, is shown in Fig. 16. Since the quality-controlled observations used in the verification are located at the background GFS pressure levels, which have 25 hPa spacing near the ground, this is the effective maximum vertical resolution of the figure. The figure shows the zone from 700 hPa down to 975 hPa, which is the lowest pressure bin located entirely above the surface at Fairbanks. Note that the chosen variations in physics packages have virtually no effect above approximately 850 hPa, and all the modeled temperature profiles are extremely close to the observed profile, presumably due to the impact of Grid 2 obs nudging along the boundaries of Grid 3. From about 850 hPa to 925 hPa, the models begin to diverge from the observations for the near total darkness episode; the MND configuration is about one degree C too warm, but the models with the RRTMG radiation package reduce the positive temperature bias by about a factor of two. For the partial sunlight episode all models track the observations closely down to about 900 hPa. Below 950 hPa the MRR configuration becomes the coldest of the models, and the closest to the observed profile, especially for the near total darkness episode. At these lowest levels the RRTMG sensitivity remains much greater for the near total darkness episode than for the partial sunlight episode. The MNR and QNR configurations are always virtually indistinguishable, suggesting that choice of PBL scheme has little impact on the vertical temperature structure at 975 hPa and higher elevations.
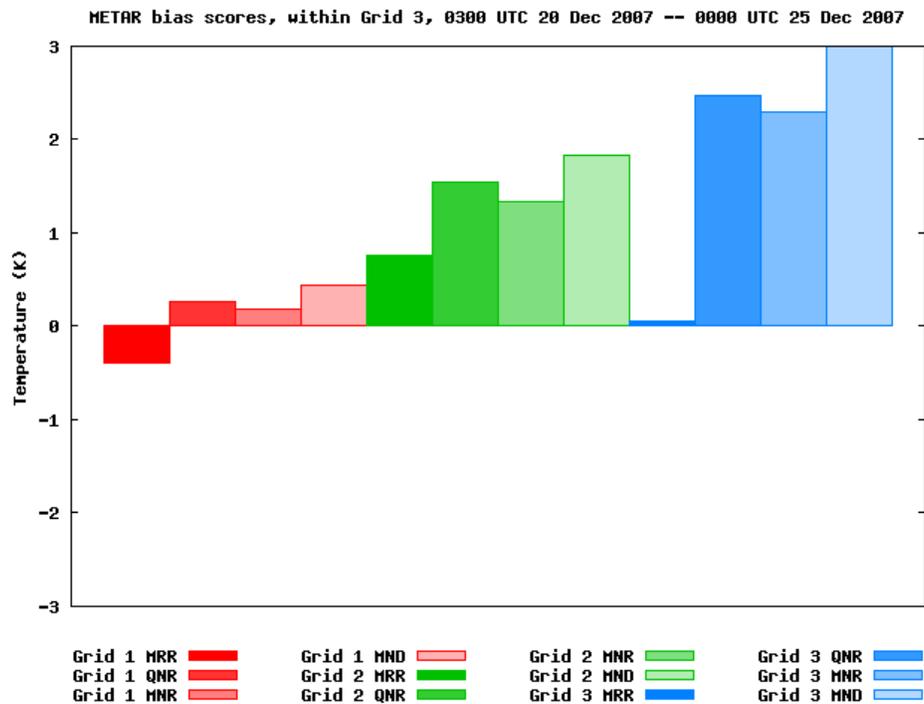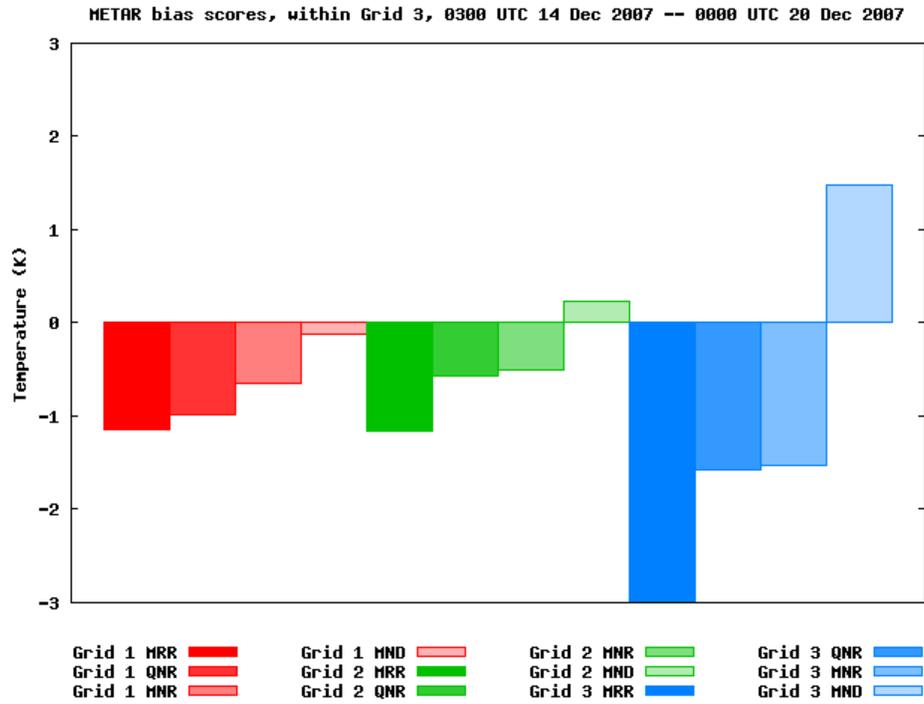
Fig. 14: Surface METAR bias scores for temperature during the near total darkness episode within the 14-20 Dec period (top) and 20-25 Dec period (bottom). Otherwise, same as Fig. 11.
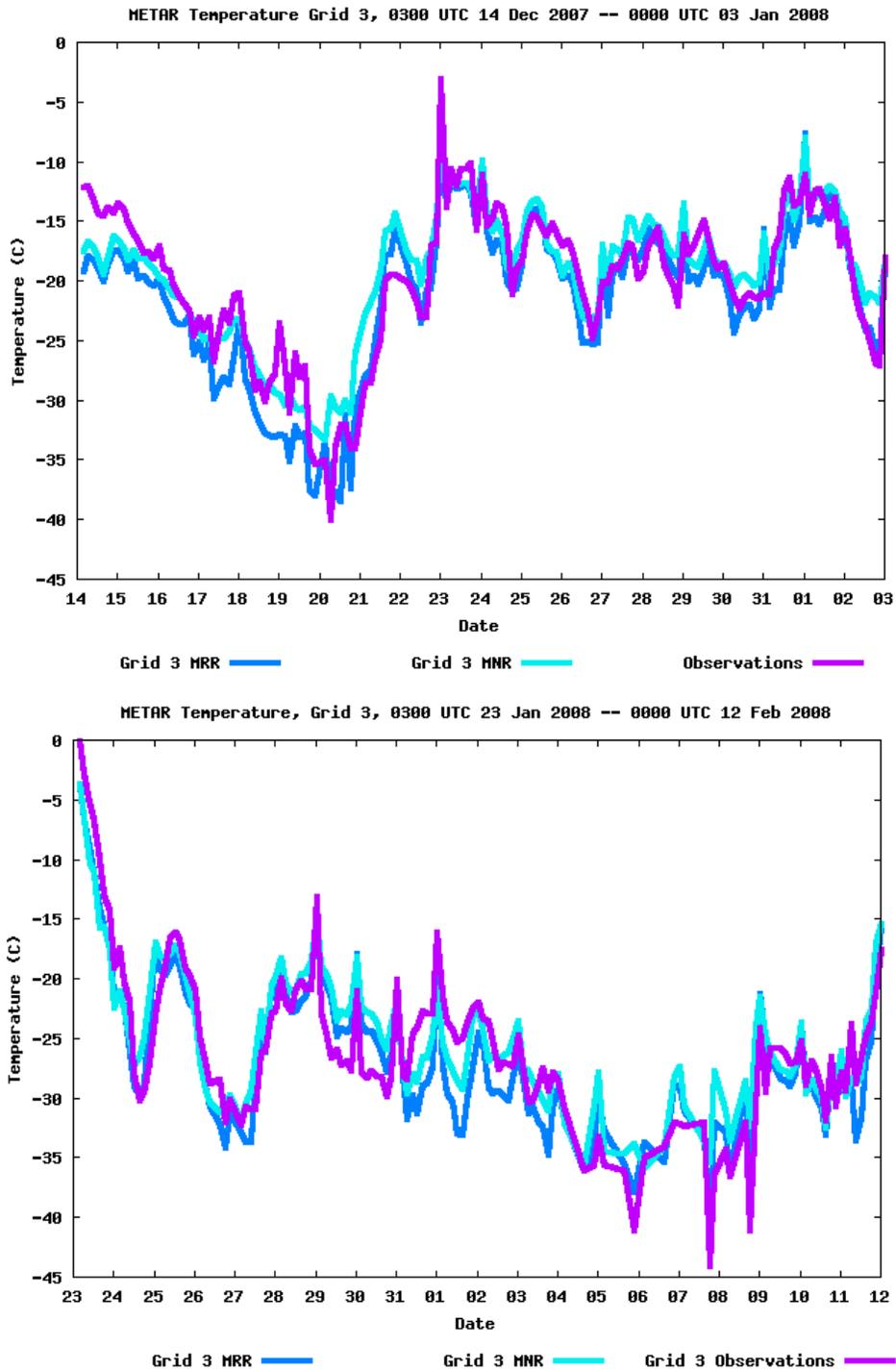
Fig. 15: Time series of temperature for near total darkness episode (top) and partial sunlight episode (bottom), averaged over the sites of all the surface METAR stations within Grid 3. Dark blue indicates value within Grid 3 from experiment MRR; light blue indicates value within Grid 3 from experiment MNR; purple indicates observed METAR value.
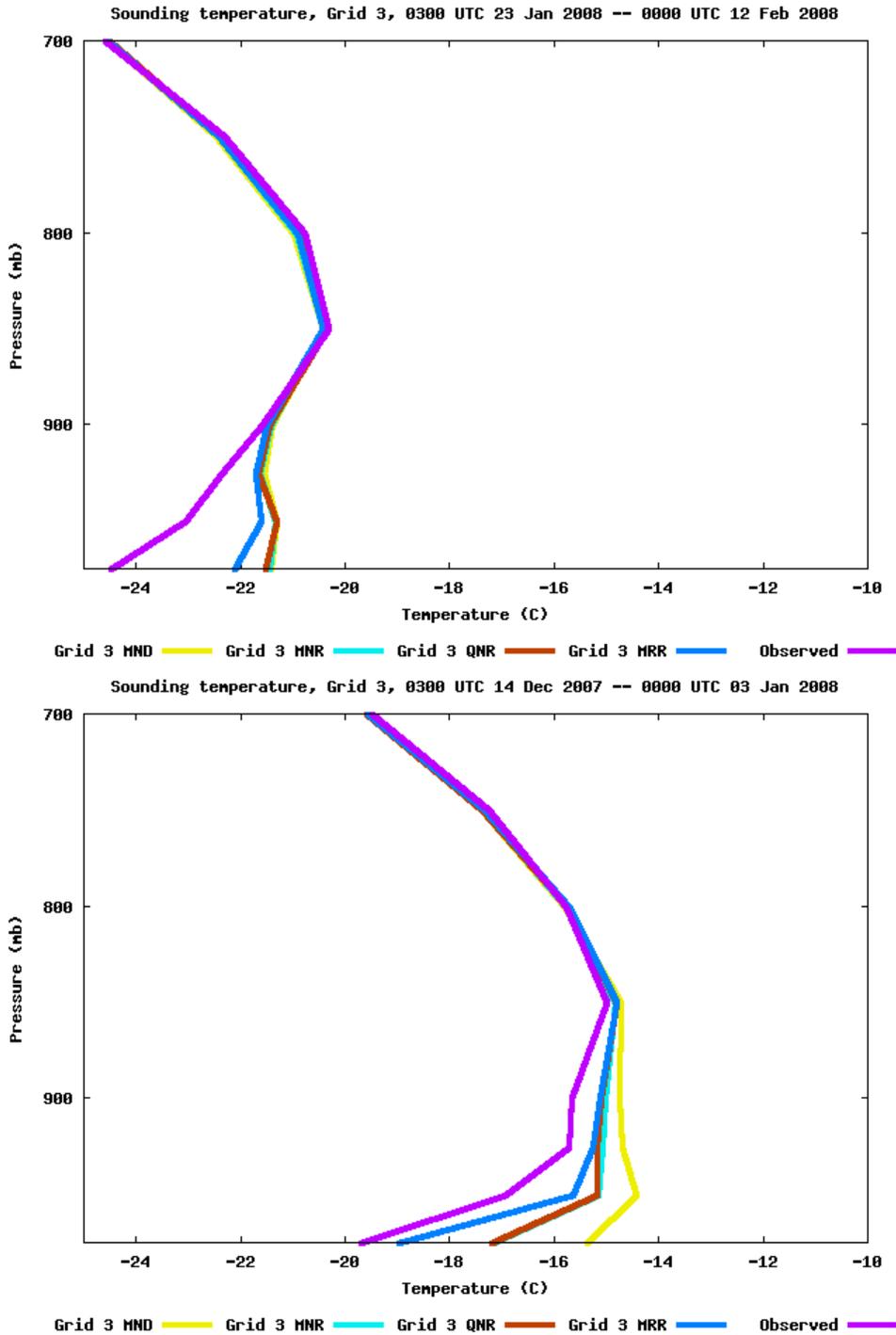
27

Fig. 16: Time-averaged vertical profile of Fairbanks sounding (PAFA) temperatures for near total darkness episode (top) and partial sunlight episode (bottom). Dark blue indicates value from experiment MRR; brown indicates value from experiment QNR; light blue indicates value from experiment MNR; yellow indicates value from experiment MND; purple indicates observed sounding value.

28

In order to learn more about the behavior of the different models, we examined the time series of the reported surface-level temperature of the raw Fairbanks sounding in comparison to the lowest-level model values at that location. When we compare the time series for the period surrounding the coldest temperatures of each episode (Fig. 17), an obvious diurnal trend appears in the partial sunlight episode observations during the coldest period from about Day 12 to Day 19 (4 – 11 Feb 2008). Little diurnal trend appears in the observations earlier in the episode; by contrast, the models all have a significant diurnal trend in surface temperature throughout the partial sunlight episode. The model diurnal amplitude during the 4-11 Feb 2008 period for the experiments other than MRR is approximately consistent with the observed amplitude, but the temperature values are consistently about 7°C too warm during this period. The MRR diurnal amplitude is somewhat larger than the others, such that it is similar to the other models for the warmer 0000 UTC times, but is much closer to the observations for the colder 1200 UTC times. For the period of rapidly falling temperatures immediately prior to 4-11 Feb the MRR experiment remains colder than the other models at the 1200 UTC times. In this case, the model 0000 UTC soundings are close to the observations, so the presence of a diurnal tendency in the model but not the observations during this time causes the 1200 UTC model soundings to be too cold, and the MRR sounding to be very cold.

During the near total darkness episode there is little diurnal variation in either the model or the observations. However, we again see the feature that when the temperatures are rapidly decreasing below -30°C, the MRR configuration has a substantial cold bias; once the coldest temperatures are achieved, however, the MRR configuration is better able to capture the low temperatures than the other models.

Finally, in order to gain as much insight as possible into the model-predicted PBL structure near the surface during the coldest episodes, we performed an alternate verification procedure using the raw Fairbanks sounding for the last ten days of the partial sunlight episode (2 – 12 Feb 2008). Instead of interpolating the model sounding to 25 hPa increments of the observed sounding, we interpolated the raw sounding to each WRF model level using some basic assumptions. (The WRF model levels are specified in terms of $(p - p_T)/(p_s - p_T)$ where p_T is the specified model top pressure, p_s is the surface pressure, and all pressures are the dry hydrostatic components; here we converted each WRF level to a pressure in the Fairbanks sounding using the observed surface pressure and the assumption that the actual pressure is approximately the dry hydrostatic pressure; the temperature at the resultant pressure was found by log-pressure interpolation. Finally, the physical height for each WRF level in the base state over ocean was used to determine the abscissa coordinate in Fig. 18.) This procedure gives us increased vertical resolution near the surface, where the model levels are much closer than 25 hPa (i.e., roughly 250 m in physical distance) apart. A plot of the temperature structure (Fig. 18) shows the same general trends as in the 25-hPa plot of Fig. 16. The two simulations using the Noah scheme (QNR and MNR) are similar, while the simulation using the RUC LSM (MRR) is consistently colder in the lowest 500 m. However, all simulations have a warm bias in the lowest 700 m. Though in the lowest 100 m the models have average vertical temperature gradients as large as, or even larger than, those in the observations, the vertical temperature gradients comprising the inversions in the observations extend to a much

greater altitude, consistent with the significantly greater temperature differences between the surface and the 1-km level found in the observations.

In summary, it appears that during periods of near total darkness and the cold, dry, calm conditions characteristic of high fine particulate concentrations, all models possess a low-level warm bias, but the bias is minimized and the statistics are generally the best in the MRR (MYJ PBL / RUC LSM / RRTMG radiation) experiment. The reason for the improved statistics in MRR for the extremely cold episodes is not precisely known at present, but it is probably related to some combination of its potentially multi-level snow model (which can serve to reduce the heat flux from the ground to the atmosphere) and the presence of a ground surface 'skin' layer in the RUC LSM (which has no thermal inertia itself and could decrease the time needed for the ground and the adjacent atmosphere to respond to a negative heat budget). A few caveats are in order, however. During the period of decreasing temperatures immediately preceding temperatures below approximately -35°C, the MRR configuration is still colder than the other models, but for these periods MRR tends to exacerbate an already cold model bias instead of improve a warm model bias. Since the observations for these falling temperature periods tend to show fairly continuous frozen precipitation (in contrast to the coldest temperature periods which tend to have ice fog and no precipitation), it is possible that all the model configurations have difficulties with modeling the microphysics/radiation interaction. For example, if the radiation scheme is not taking into account the presence of ice crystals when they exist in the actual atmosphere, the absence of their radiative heating effect on the surface during these extremely cold conditions could be significant.   Another caveat is that when partial sunlight is present, MRR tends to warm more rapidly than the other models, and all models tend to have substantial warm biases in these conditions.
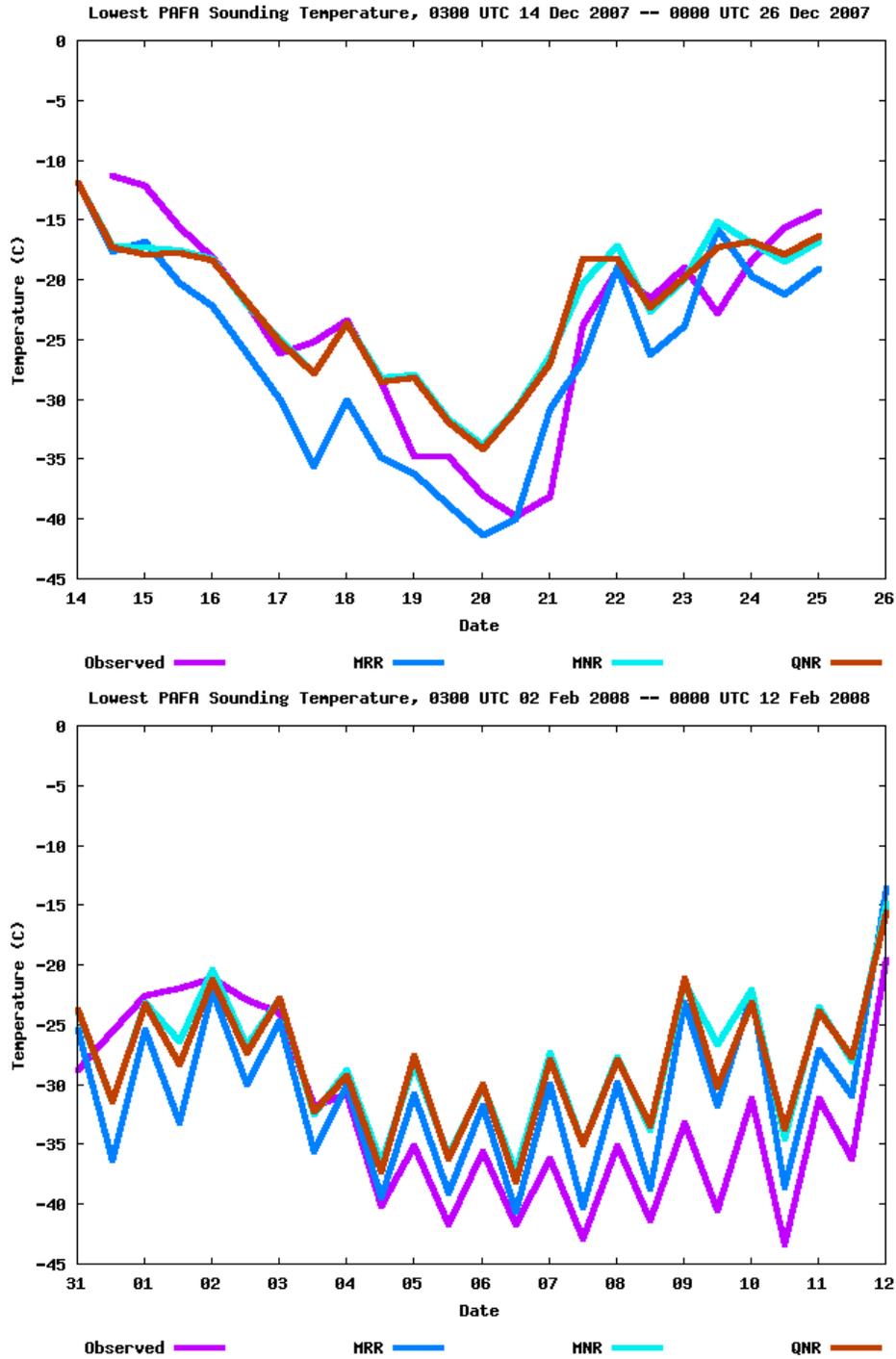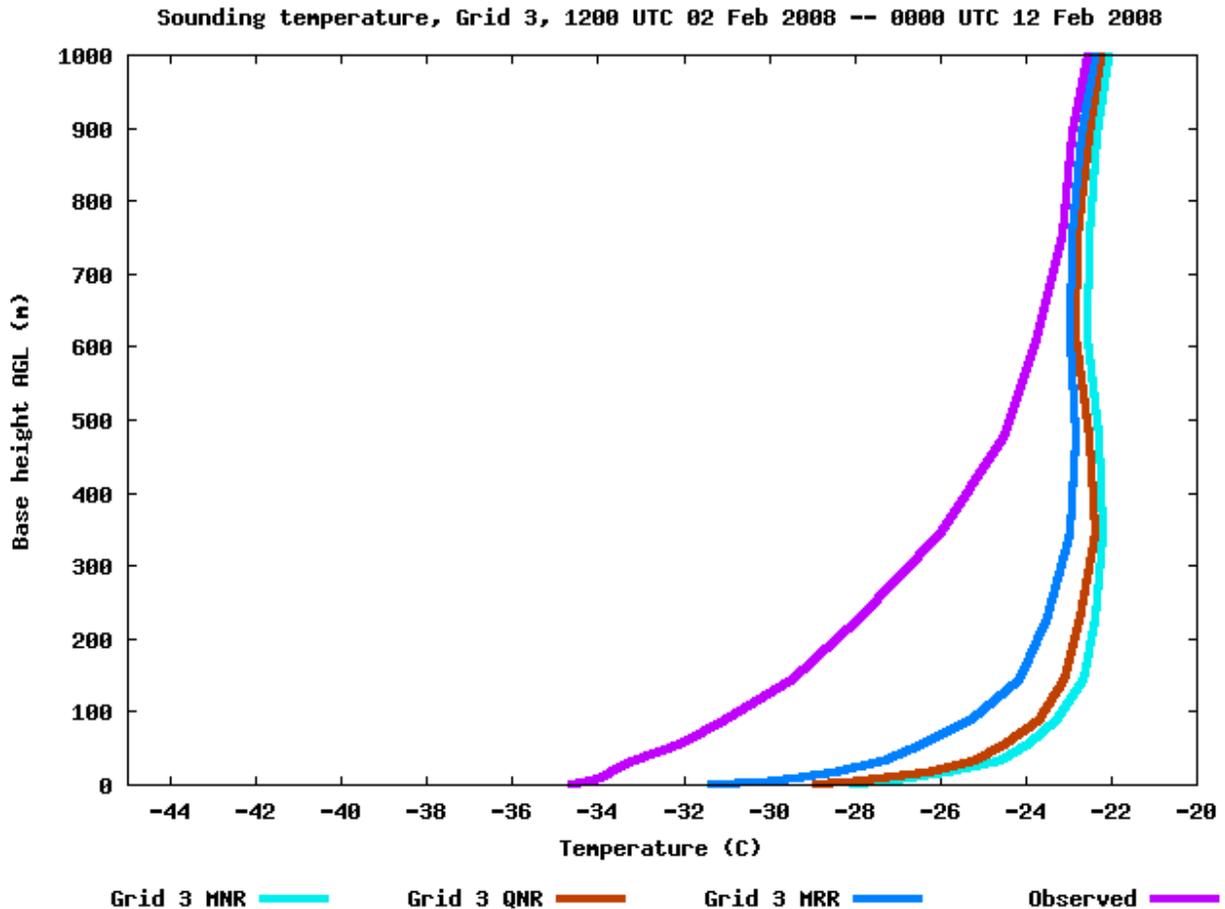
Fig. 17: Time series of raw Fairbanks surface-level reported sounding temperatures (PAFA) for 14-26 Dec 2007 period of near total darkness episode (top), and 02-12 Feb 2008 period of partial sunlight episode (bottom). Colors are same as in Fig. 16.

Fig. 18: Time-averaged vertical profile of Fairbanks sounding (PAFA) temperatures for 02-12 Feb 2008 period during partial sunlight episode, where observations are interpolated to WRF vertical levels as described in text. Dark blue indicates value from experiment MRR; brown indicates value from experiment QNR; light blue indicates value from experiment MNR; purple indicates observed sounding value.

5. SELECTION OF PREFERRED PHYSICS CONFIGURATION AND FINAL DYNAMIC-ANALYSIS SIMULATION

Based on the results of the physics sensitivity test, we concluded that the physics suite contained in experiment MRR (MYJ PBL, RUC LSM, and RRTMG radiation) was the best one to be used to simulate the two episodes. The high exceedance events that are of importance occur during the coldest temperature periods when the RUC LSM appeared to perform the best. However, the tendency of the MRR suite to produce significant negative temperature biases during the falling temperature periods must be noted.

We thus concluded that we should perform an additional dynamic-analysis simulation with the MRR physics package but with Grid 3 obs nudging in order to reduce the noted temperature biases and generate the best atmospheric analysis. Because the MRR Grid 2 statistics, in the presence of obs

nudging, were almost always quite good, we were optimistic that any systematic biases present in the MRR simulation on Grid 3 would be greatly alleviated through obs nudging. As noted above, however, we did not nudge the wind fields from surface data on Grid 3, whose influence is below ~150 m (see Fig. 8), because of concerns that the local topographic drainage flows generated by the model in the topography around Fairbanks may be smoothed out by the FDDA procedure. However, we did retain nudging of wind fields on Grid 3 for observations above the surface (i.e., from the Fairbanks sounding).

The initial specifications of the parameters used for the Grid 3 obs nudging simulation are listed within the parentheses of Table 2. They closely correspond to values on the other grids. However, the value of RINXY (a horizontal radius of influence) was decreased on Grid 3 from a value of 100 km to 75 km. This value was determined by performing a temporal correlation of the Grid 3 temperature innovations within the MRR no-FDDA simulation at the location of the METAR stations, and estimating the characteristic horizontal distance at which the Grid 3 METAR observational-based surface innovations were correlated for (see Fig. 19). The surface pressure difference parameter used in the horizontal weighting function in complex terrain (henceforth Δp_d) was also reduced from 75 hPa to 37.5 hPa based on the results of the correlation analysis (e.g., note the relationship between correlation value and the elevation difference labels in Fig. 19). This parameter controls how far the influence of a surface observation may spread along topography as the surface pressure varies from that at the obs site; our results suggested that some station pairs close in horizontal distance but with different vertical elevations were much less correlated than similar stations with little terrain difference.

An additional complication derives from the fact that the WRF method of reducing the weight of surface observations based on Δp_d is different from the MM5 method defined in Stauffer and Seaman (1994). In default WRF, if there is a difference between the model surface pressure at the location of a surface observation, p_b, and the model surface pressure at a grid point in question, p, the weight of the surface observation is reduced by a factor w given by:

$$W = \max\left[0.0, \ 1.0 - \frac{|p - p_b|}{\Delta p_d} \left[\frac{r_0^2 - r^2}{r_0^2 + r^2} \right] \right], \tag{1}$$

where r is the horizontal distance between the grid point and the observation, and r_0 is the surface radius of influence parameter (RINXY in Table 2). In MM5, on the other hand, the surface pressure difference is used to artificially increase the horizontal radius of influence parameter, according to:

$$w = \frac{r_0^2 - \left[r + r_0 \dfrac{|p - p_b|}{\Delta p_d} \right]^2}{r_0^2 + \left[r + r_0 \dfrac{|p - p_b|}{\Delta p_d} \right]^2} .$$

(2)

Though the two functions are often similar, the WRF function tends to be more horizontally isotropic and less sensitive to terrain features, as well as generally nonzero over greater horizontal differences. (The WRF method will give nonzero weights to surface observations unless either $|p-p_b|$ exceeds Δp_d or r exceeds r_0, whereas the MM5 method can give a zero weight even if neither criterion is met because the terrain difference increases the effective distance from observation to grid point.)   In the final Grid 3 FDDA simulations used here, the MM5 method for surface pressure difference weighting was used.
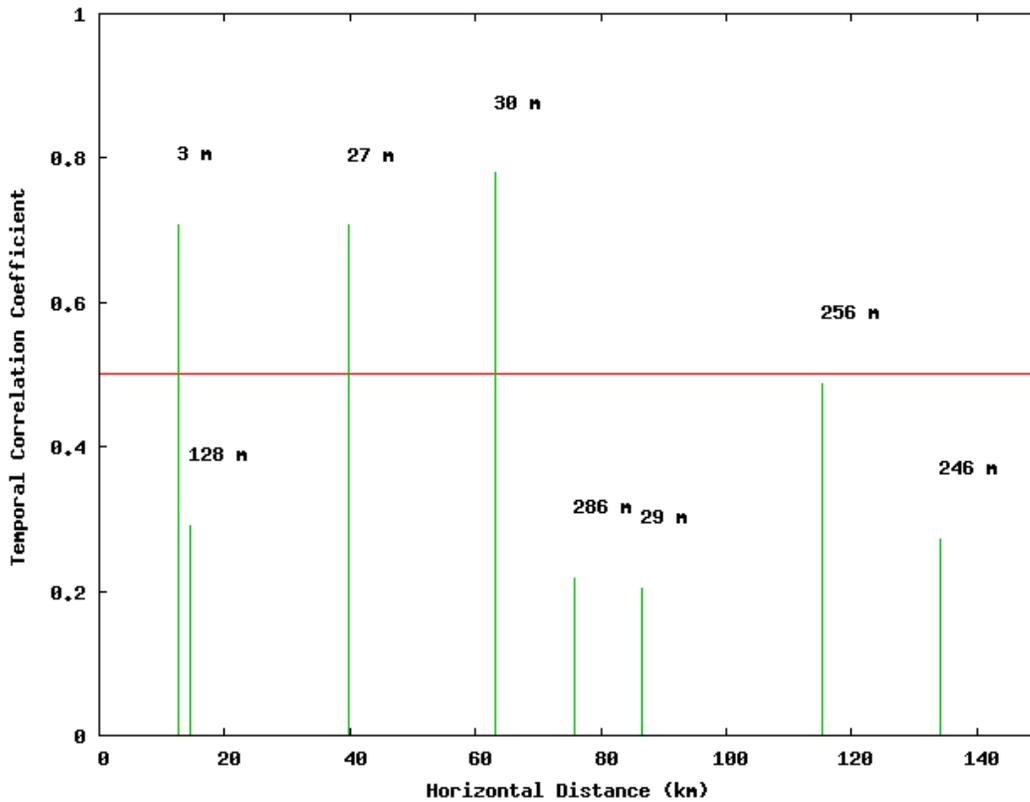

Fig. 19: Temporal correlation coefficients vs. horizontal separation distance between various pairs of surface METAR stations located within Grid 3 (green). Red line indicates a temporal correlation coefficient of 0.5. Numerical labels indicate elevation distance between stations in meters.

The value of TWINDO (Table 2), the obs nudging time window half-period defining the temporal influence of an innovation (Stauffer and Seaman 1994), should also be considered. Ideally this parameter would be a function of height and decrease in value towards the surface, to take into account the shorter temporal correlation time scales for surface data reflecting surface processes. Although this capability will be present in WRF version 3.2, in the version 3.1 that we used for this study, it is simply a constant (though it may vary with grid). Our experience suggests that the value chosen, 2.0 hours, is generally best for the assimilation of sounding data, but may be somewhat too large for the surface (Schroeder et al. 2006). For the final version of the Grid 3 FDDA simulations, we manually encoded the Penn State MM5 method used in WRF version 3.2 so that the effective value of TWINDO was 2.0 hours above the surface, but 1.0 hours at the surface.

Finally, two additional modifications were made to the default WRF version 3.1 FDDA procedure. In the default procedure the surface level observation of a sounding is treated differently than a surface observation. Specifically, a surface observation is assumed to be applicable to the lowest model level at the horizontal location of the observation, whereas all sounding observations including one at the surface level are assumed to be applicable at the vertical model location with the same pressure as the pressure of the observation. So a sounding surface level observation will not necessarily be placed at the lowest model level if the model surface pressure is not the same as the observed surface pressure. Also, the surface pressure difference is used to reduce the weight of a surface observation at remote horizontal grid points , but not the weight of a surface-level sounding observation. This inconsistent treatment becomes more of an issue in station-poor regions such as that of the Grid 3 used in this study, where the relative influence of the Fairbanks sounding to all Grid 3 METAR stations may be quite significant. In the final Grid 3 FDDA simulations, the code was modified to remove the surface-level observation from the rest of the sounding and treat it as a separate surface observation. Furthermore, to reflect the Penn State MM5 method, the Δp_d weighting function was applied to soundings as a unit, in addition to surface observations.

Figures 20-22 show the vertical profiles of RMSEs verified against the Fairbanks sounding for a series of trial simulations of the first six days of the near total darkness episode (14-20 Dec. 2007) using the MRR physics suite but different variations of the Grid 3 obs nudging procedure. First, the benefit of Grid 3 obs nudging is immediately apparent, and Fig. 21 shows in particular that the simulations with retained Grid 3 wind obs nudging above the near-surface layer have substantially reduced wind speed RMSE scores in comparison with the two simulations that don't. This helps justify our proposed procedure of retaining Grid 3 wind obs nudging above the near-surface layer but turning it off within the near-surface layer to allow the model to generate its own topographic flows. Second, for the non-wind fields shown in Fig. 20 and Fig. 22 we see that the TWINDO = 2.0 hours statistics tend to be somewhat better than the TWINDO = 0.45 hours statistics, in agreement with our past experience. The proposed Grid 3 obs nudging procedure, including, among other modifications, using TWINDO = 2.0 hours above the surface but TWINDO = 1.0 hours at the surface, produces results quite similar to the TWINDO = 2.0 hours simulation.

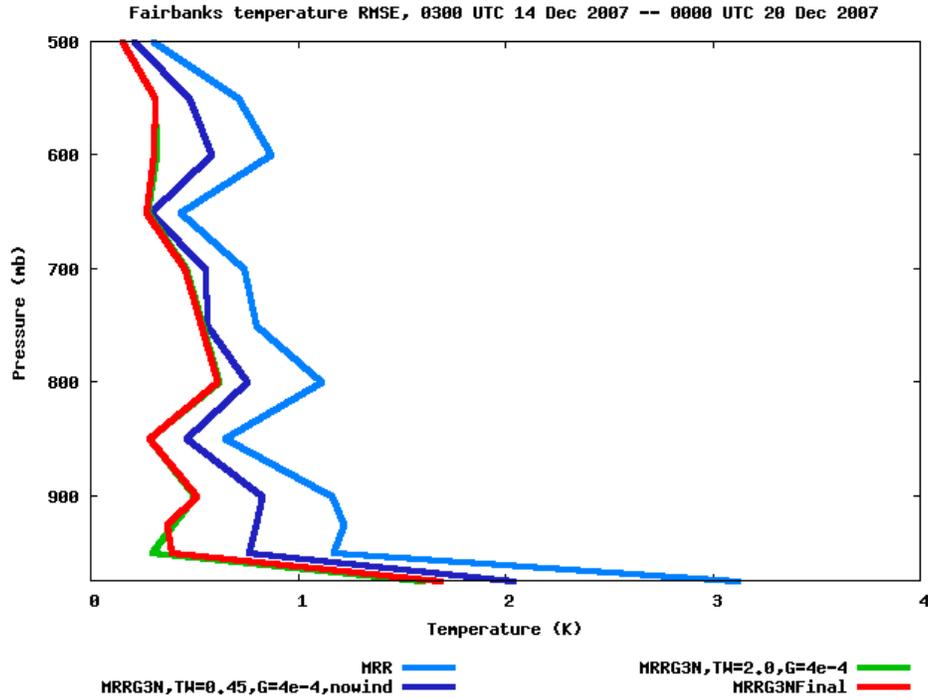Fairbanks temperature RMSE, 0300 UTC 14 Dec 2007 -- 0000 UTC 20 Dec 2007

Fig. 20: Time-averaged vertical profile of Fairbanks sounding (PAFA) temperature RMSE scores for 14-20 Dec 2007 period of near total darkness episode. Blue indicates the value from experiment MRR; violet indicates the value from MRR experiment using default Grid 3 obs nudging with TWINDO = 0.45 hours and no wind nudging; green indicates the value from MRR experiment using default Grid 3 obs nudging with TWINDO = 2.0 hours and nudging of wind above the near-surface layer only; red indicates the value from MRR experiment using the final version of Grid 3 obs nudging with the modifications as described in the text.
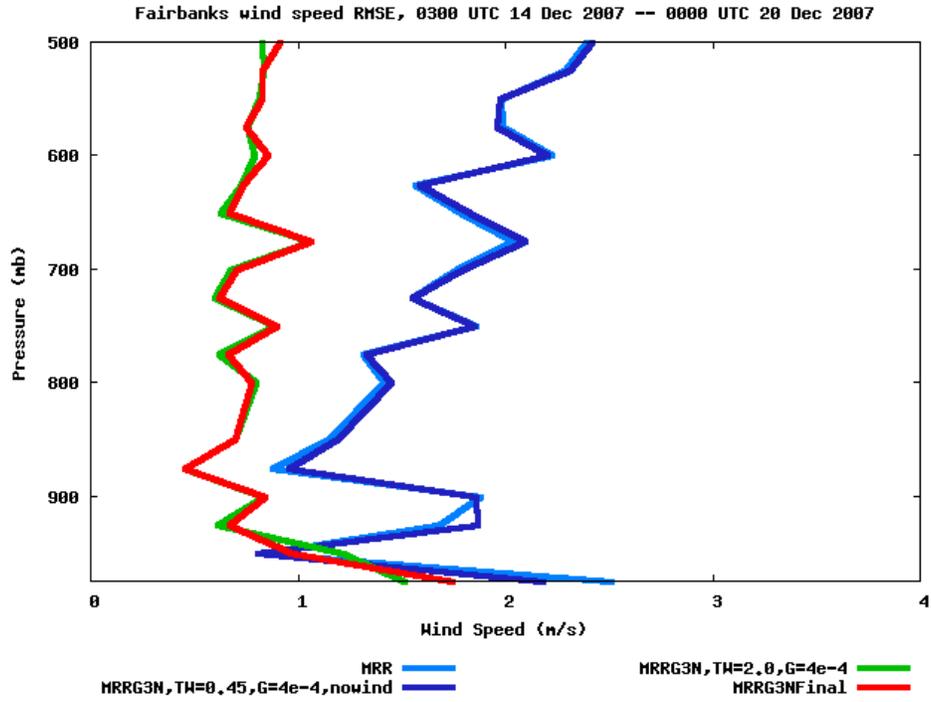
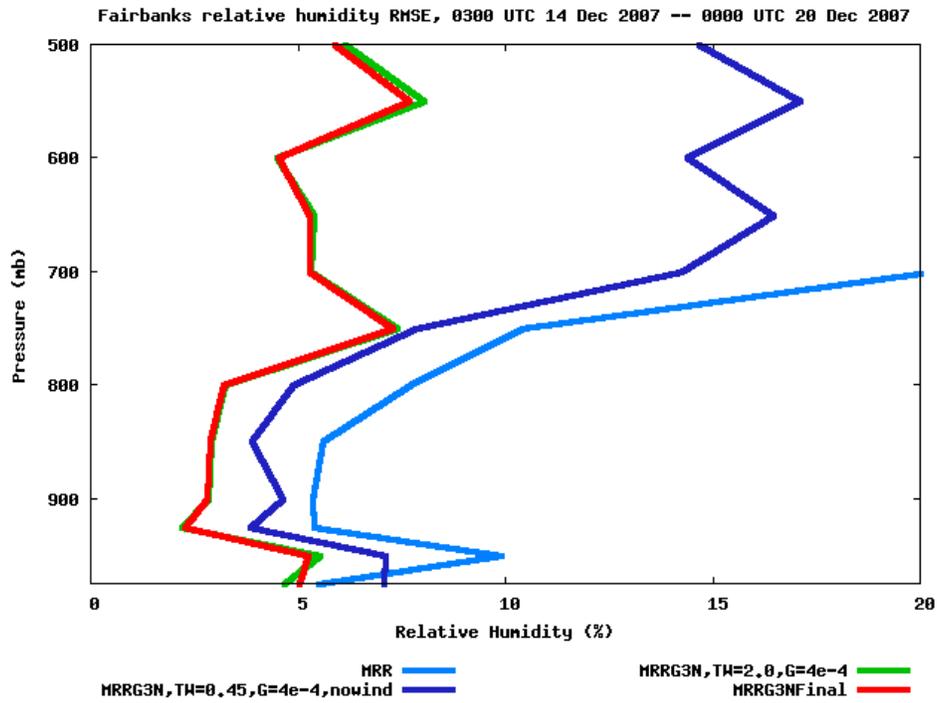*Fig. 21: Same as Fig.20, but for wind speed RMSEs.*

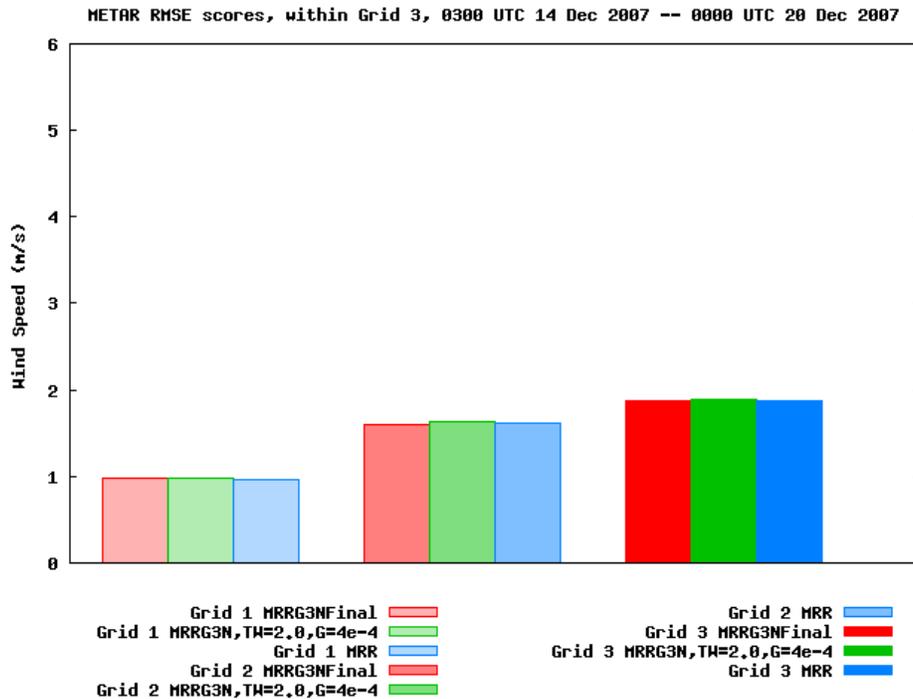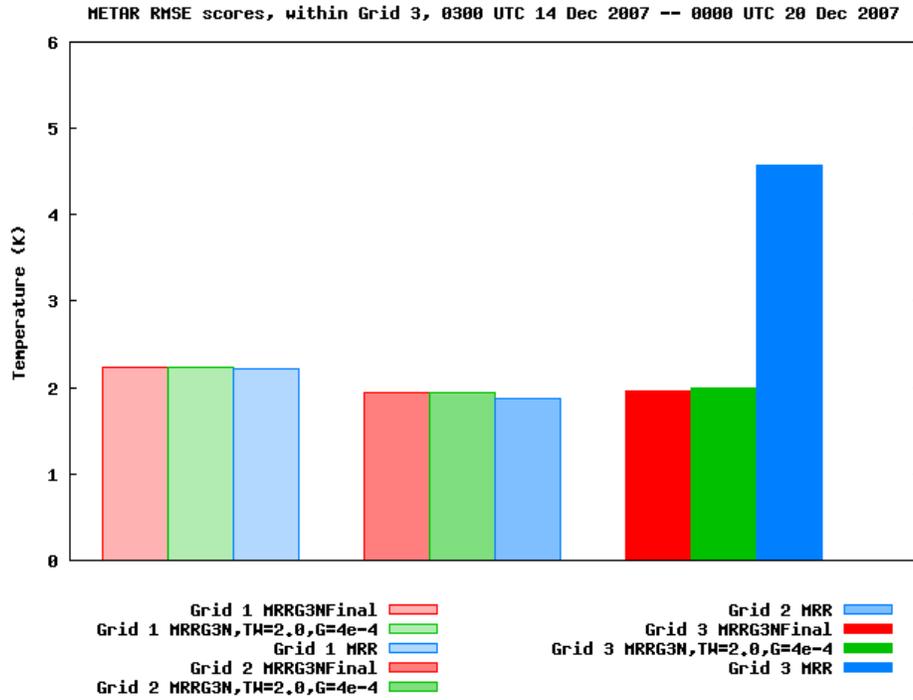*Fig. 22: Same as Fig. 20, but for relative humidity RMSEs.*

37

Fig. 23: Surface METAR RMSE scores for during 14-20 Dec 2007 period of near total darkness for temperature (top) and wind speed (bottom). Blue indicates value from experiment MRR; green indicates value from experiment MRR using default Grid 3 obs nudging with TWINDO = 2.0 hours and nudging of wind above the near-surface layer only (i.e., MRRG3N,TW=2.0,G=4e-4); red indicates the value from MRR experiment using the final version of Grid 3 obs nudging with the modifications as described in the text (i.e., MRRG3NFinal).

Fig. 23 shows RMSE statistics for the sample period for the surface METAR stations within Grid 3. The lightest, medium, and darkest shades in the histogram plot correspond to the dark blue, brown, and yellow curves in the vertical profile plots. In all cases the improvement of the MRR temperature RMSE scores from the Grid 3 obs nudging is quite dramatic, and shows the utility of our dynamic analysis approach. The fact that some of our modified obs nudging procedures carried over to all grids caused the Grid 1 and Grid 2 results to change from those of the MRR experiment, but the magnitudes of the changes are small. Wind speed statistics for the surface METARs show little sensitivity to the presence of either Grid 3 obs nudging of temperature and humidity, or Grid 3 obs nudging of winds above the near-surface layer. The proposed Grid 3 obs nudging procedure produces only slight differences from those of the more standardized Grid 3 obs nudging procedure shown, but to the extent there are differences they are generally slight improvements.

In summary, the use of our proposed modified Grid 3 obs nudging procedure, at least for this six-day test period, produces the desired effect of greatly improving the surface temperature statistics without significantly degrading the other statistics, and is also consistent with our past experience as to the preferred specification of obs nudging parameters. Therefore, we proceded to perform the final dynamic-analysis simulations in their entirety using the proposed Grid 3 obs nudging procedure.

Figures 24-26 show the overall statistics for the final dynamic-analysis Grid 3 obs nudging simulation for the entire near-total darkness episode in comparison to those of the non-Grid 3 obs nudging simulation MRR. The final temperature biases in comparison to the surface METARs are below 0.5°C in magnitude, with RMSE errors 2-3°C. The temperature RMSE errors decrease below 1°C above 900 hPa. Wind speed biases are under 1 m s^{-1} at the surface, while RMSE errors are on the order of 2 m s^{-1} throughout the lower troposphere. Qualitatively, the statistics for the final partial sunlight episode (Figs. 27-29) show very similar tendencies.
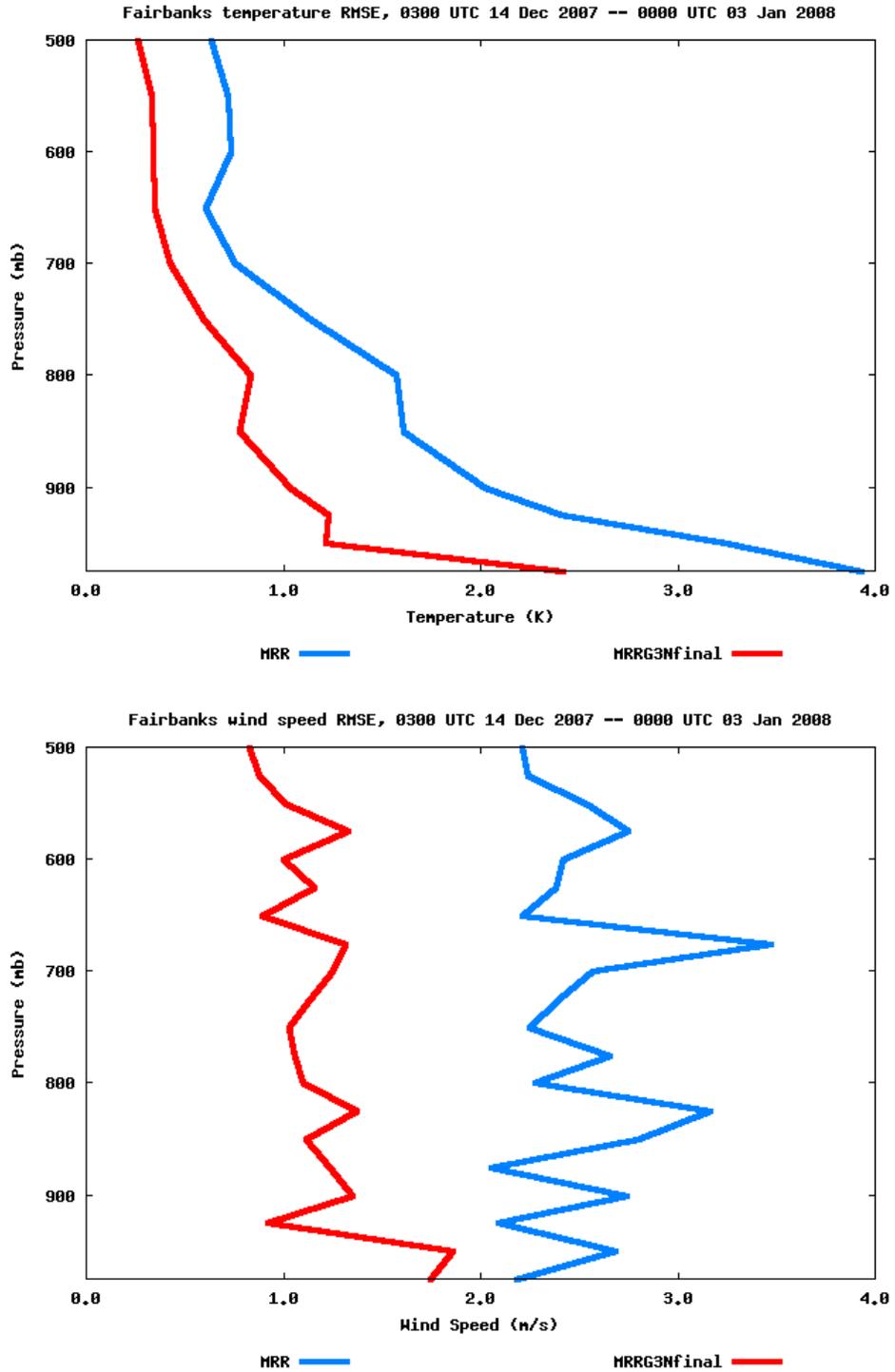
Fig. 24: Time-averaged vertical profile of Fairbanks sounding (PAFA) on Grid 3 for temperature (top) and wind speed (bottom) for 14 Dec 2007—03 Jan 2008 near total darkness episode. Blue indicates value from experiment MRR; red indicates value from final dynamic-analysis MRR simulation using Grid 3 obs nudging.
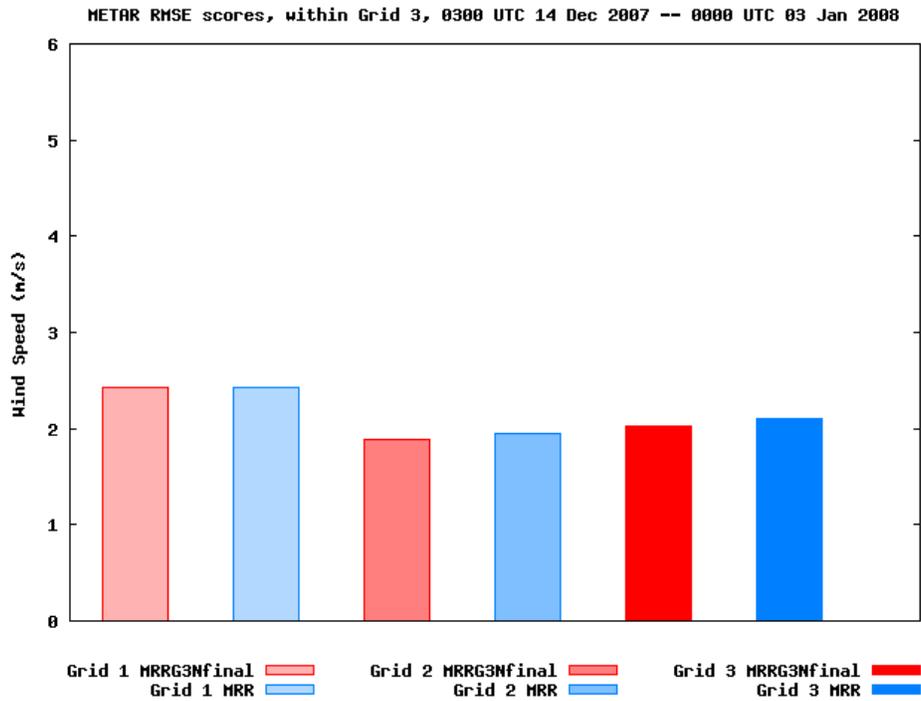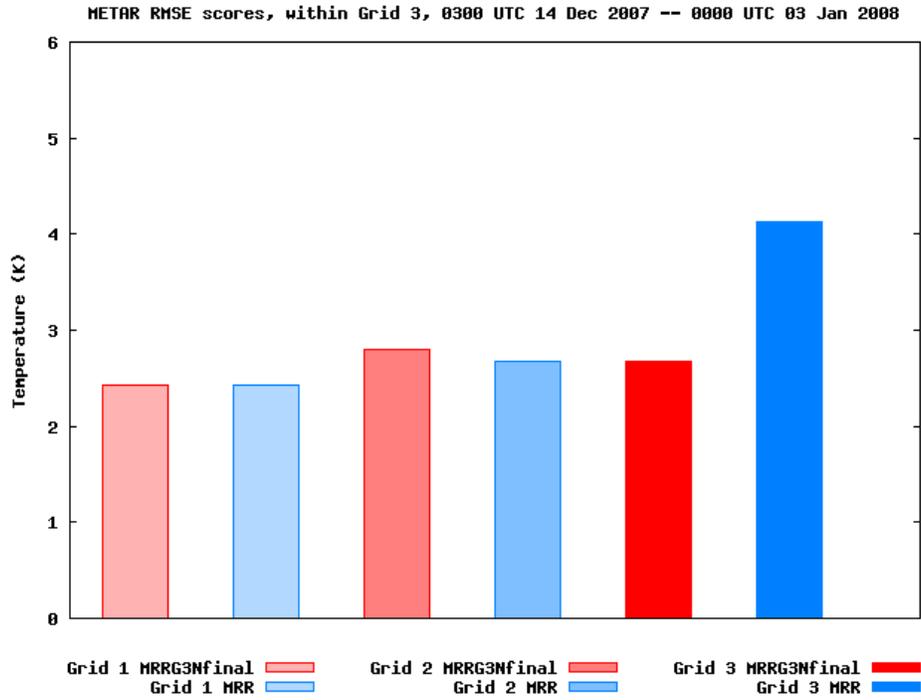
40

Fig. 25: Surface METAR RMSE scores during 14 Dec 2007—03 Jan 2008 near total darkness episode for temperature (top) and wind speed (bottom). Blue indicates value from experiment MRR; red indicates value from final dynamic-analysis MRR simulation using Grid 3 obs nudging.

METAR bias scores, within Grid 3, 0300 UTC 14 Dec 2007 -- 0000 UTC 03 Jan 2008

Grid 1 MRRG3Nfinal   Grid 2 MRRG3Nfinal   Grid 3 MRRG3Nfinal
Grid 1 MRR   Grid 2 MRR   Grid 3 MRR

METAR bias scores, within Grid 3, 0300 UTC 14 Dec 2007 -- 0000 UTC 03 Jan 2008

Grid 1 MRRG3Nfinal   Grid 2 MRRG3Nfinal   Grid 3 MRRG3Nfinal
Grid 1 MRR   Grid 2 MRR   Grid 3 MRR

Fig. 26: Same as Fig. 25, but for bias errors.

42

Fig. 27: Time-averaged vertical profile of Fairbanks sounding (PAFA) on Grid 3 for temperature (top) and wind speed (bottom) for 23 Jan 2008—12 Feb 2008 partial sunlight episode. Blue indicates value from experiment MRR; red indicates value from final dynamic-analysis MRR simulation using Grid 3 obs nudging.

43

Fig. 28: Surface METAR RMSE scores during 23 Jan 2008—12 Feb 2008 partial sunlight episode for temperature (top) and wind speed (bottom). Blue indicates value from experiment MRR; red indicates value from final dynamic-analysis MRR simulation using Grid 3 obs nudging.
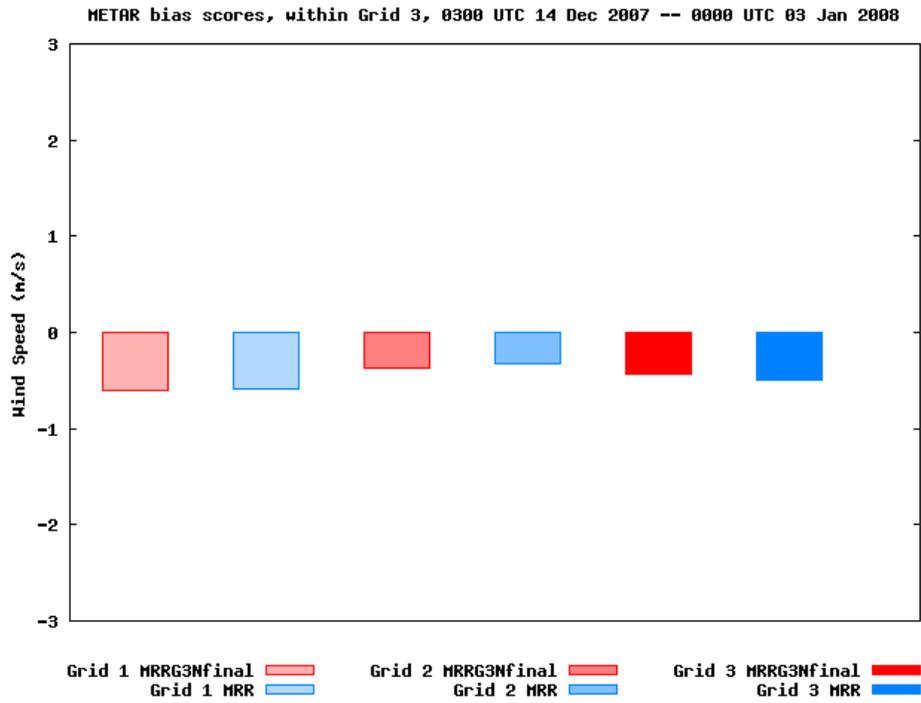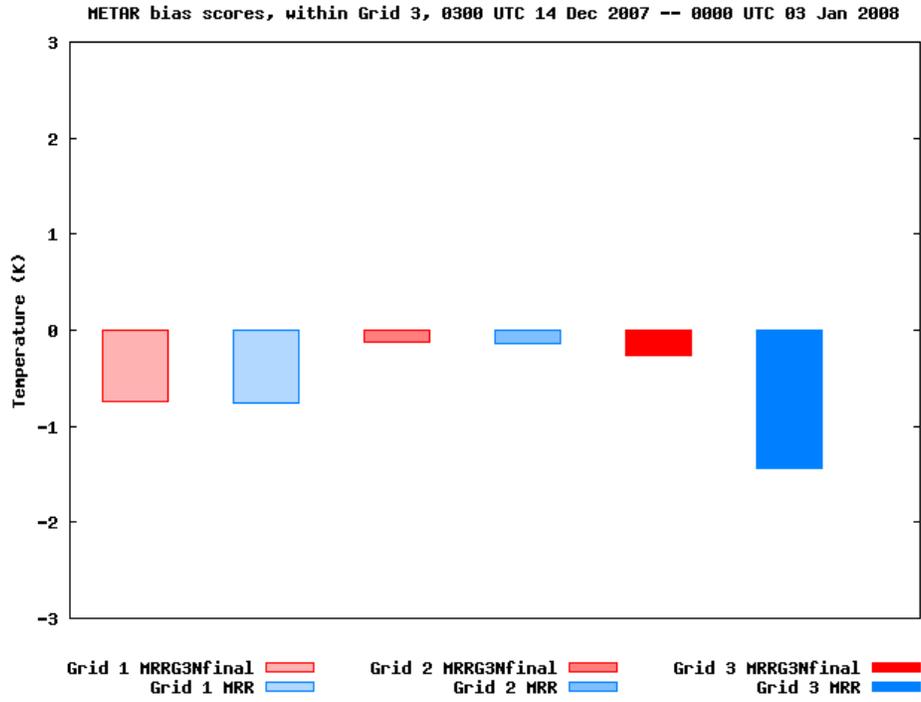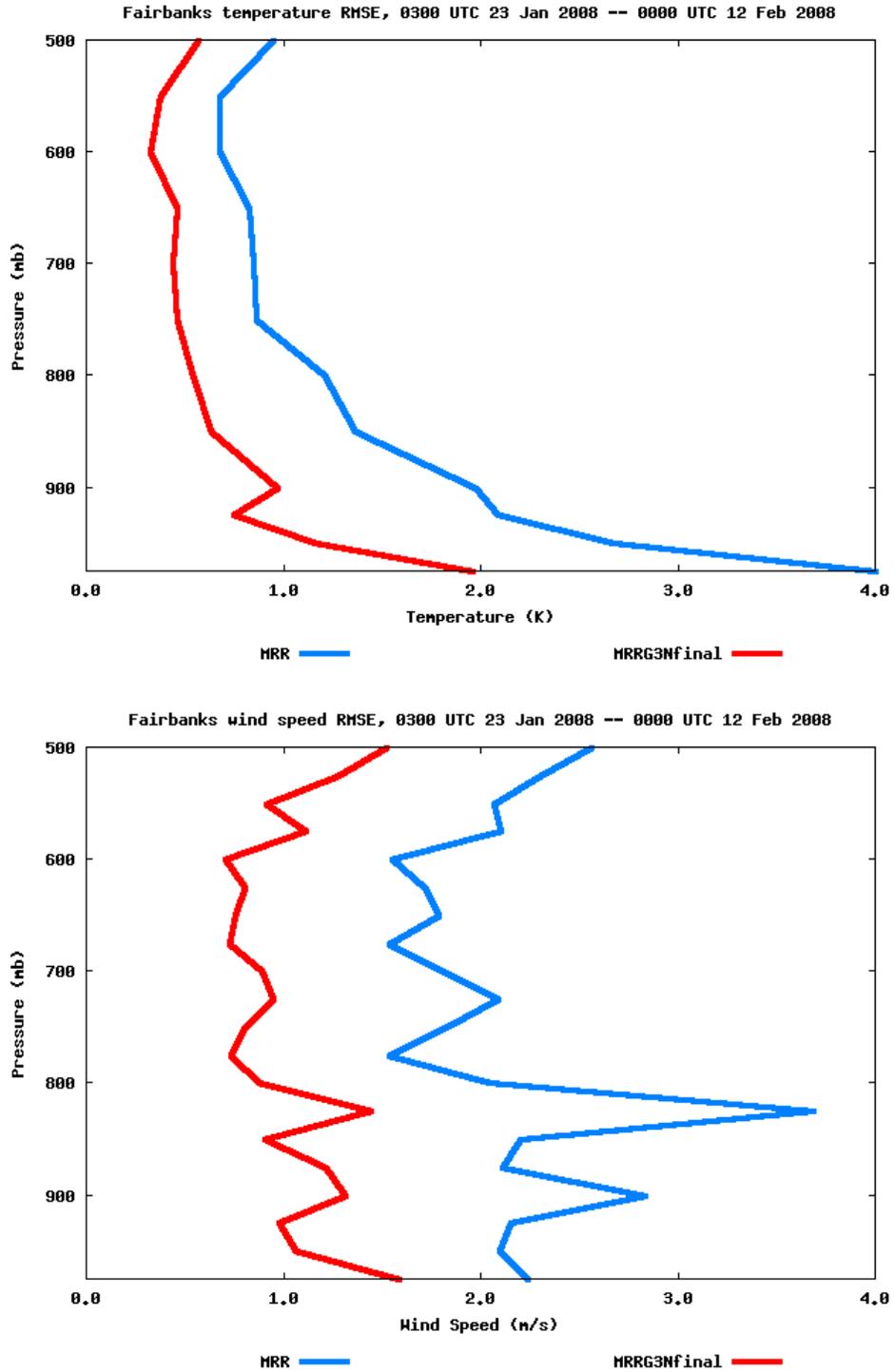
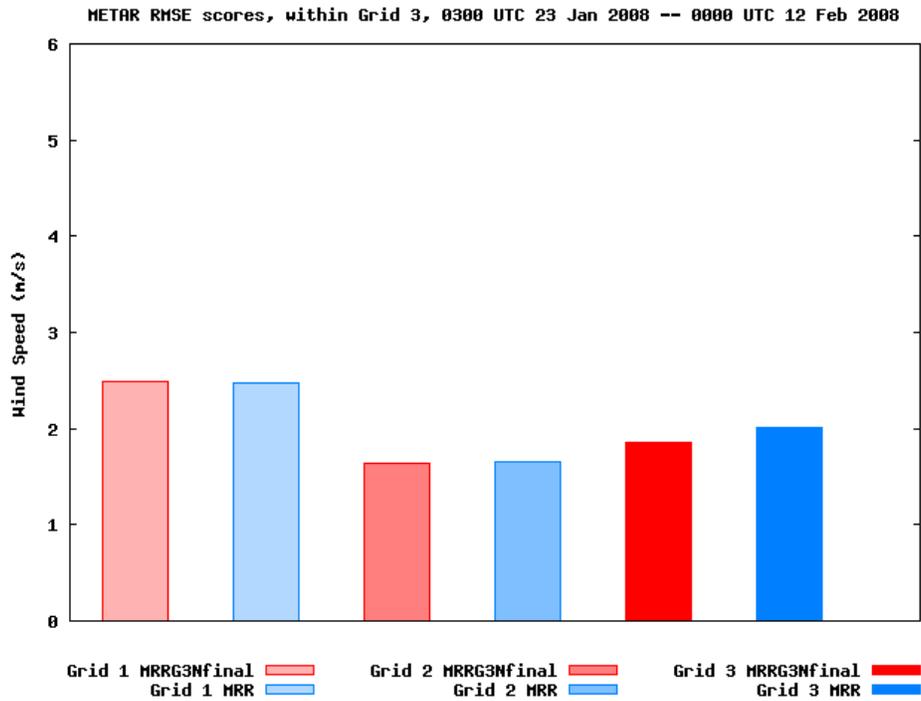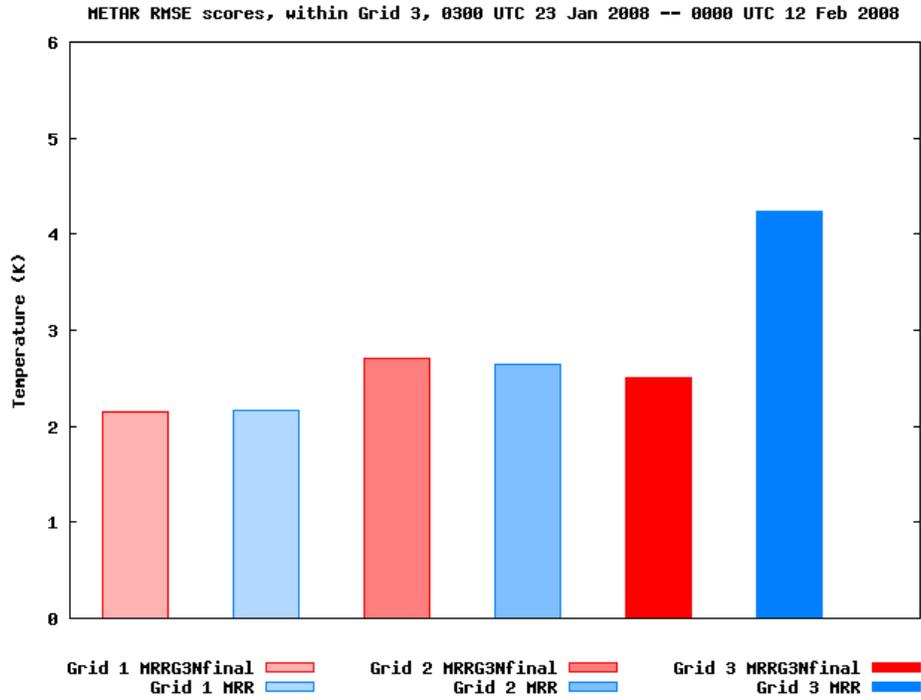METAR bias scores, within Grid 3, 0300 UTC 23 Jan 2008 -- 0000 UTC 12 Feb 2008

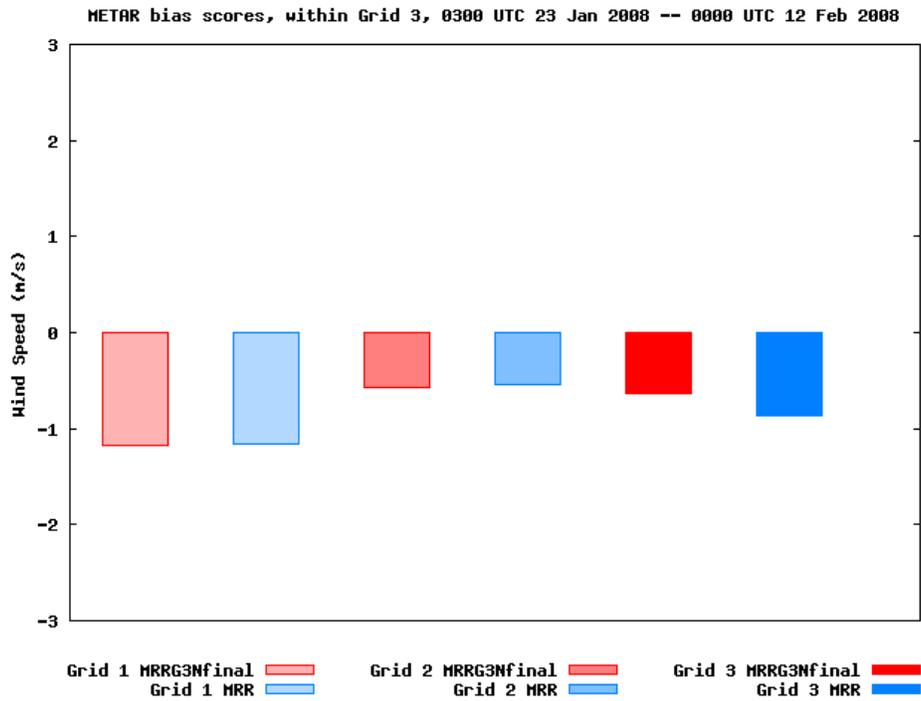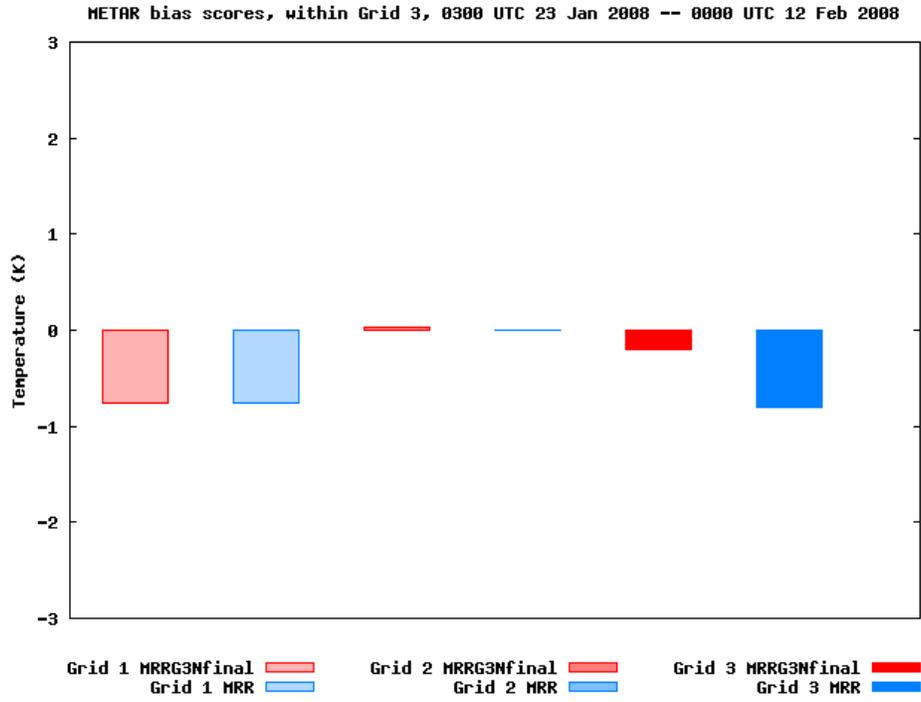METAR bias scores, within Grid 3, 0300 UTC 23 Jan 2008 -- 0000 UTC 12 Feb 2008

Fig. 29: Same as Fig. 28, but for bias errors.

6. CONCLUSIONS

6.1 Summary

The purpose of the project was to develop, adapt, and test a methodology for stable boundary layer representation (initial onset, space/time evolution, dissipation) in three-dimensional numerical models, with a specific focus on the dark, extremely cold environments such as those in the winter in the Fairbanks, AK region. A particular concern is the frequent occurrence of very high fine particular matter (PM2.5) concentrations within the stable boundary layers that form in these conditions.

Ten tasks were defined in the Statement of Work (SOW) for this project. A summary of these tasks and a brief overview of the work completed can be found in the Appendix to this report. Two twenty-day episodes were selected from the 2007-2008 winter season to study periods of extremely cold temperatures and high PM2.5 concentrations and to evaluate model performance: one in near total darkness (14 Dec 2007 – 03 Jan 2008), and the other in partial sunlight (23 Jan 2008 – 12 Feb 2008). One baseline physics configuration and three physics sensitivity experiments were performed for each episode. The physics sensitivity experiments were used to assess the impact of different planetary boundary layer (PBL) parameterizations, land surface models, and atmospheric radiation schemes on the simulations. Each simulation used three nested grids: Grid 1 (12-km horizontal grid spacing) and Grid 2 (4-km) utilized the multiscale multigrid data assimilation strategy of Stauffer and Seaman (1994) in order to ensure the model and observations remained close over the extended duration of the simulations. Grid 3 (1.3-km), centered over the Fairbanks region, did not use any direct data assimilation, and so was best-suited for quantifying the physics sensitivity; it also possesses sufficient horizontal resolution to be used by the EPA as meteorological input to chemical and air transport and dispersion models. From the different physics packages one was to be recommended to the EPA for further mesoscale modeling of the region.

The use of the three-grid configuration with a multiscale, multigrid four-dimensional data assimilation (FDDA) strategy on the outer two grids and no direct FDDA on Grid 3 consistently produced qualitatively plausible atmospheric fields throughout the variety of meteorological conditions found in the episodes, despite the relatively sparse data density. Quantitatively, the multiscale, multigrid FDDA strategy led to improved root-mean-square-error (RMSE) scores for both wind and temperature on all grids. The FDDA on the outer domains had the desired effect of improving the simulations of Grid 3 without FDDA and used for physics sensitivity tests, by providing improved lateral boundary conditions.

The best RMSE scores for the combination of both surface and sounding data required modification of the default FDDA procedure. These modifications included applying surface wind observational data to the third model vertical level instead of the lowest model level, because wind observations are normally taken at a height of 10 m which is the height of the third level in the high vertical resolution configuration used here. The influence of surface observations was also restricted to approximately the lowest 100 m, instead of to the top of the PBL, because the model-predicted PBL height in these simulations, based on the turbulent kinetic energy profile, was often found to be 1 km or higher. This

46

correction applied the surface innovation (observation minus model value) in these predominantly stable boundary layers over a much shallower layer than in the default FDDA procedure and produced improved statistical results in the lower troposphere.

All model physics combinations tended to have a positive temperature bias on Grid 3, especially during the most extremely cold periods. All of the physics sensitivity tests tended to reduce the warm bias in comparison with the selected baseline physics package. Switching from the RRTM longwave / Dudhia shortwave radiation package to the RRTMG longwave and shortwave radiation package led to significantly reduced warm biases and better RMSE statistics. RRTMG was then used in all future physics sensitivity tests. The reduced warm bias seemed to be due to the longwave component, both because of direct examination of surface fluxes in the partial sunlight case, and due to the fact that the difference was more pronounced in the near total darkness episode.

Though none of the four physics suites tested in the study was unambiguously superior to all of the others in terms of RMSE statistics, the simulation with the Rapid Update Cycle (RUC) land surface model, the Mellor-Yamada-Janjić (MYJ) PBL model, and the RRTMG radiation package was selected as the one to be recommended to EPA for modeling extremely cold SBLs and as the basis for producing the final atmospheric analysis. For both the near-total-darkness and partial sunlight episodes, the MYJ/RUC/RRTMG (henceforth MRR) physics suite had the lowest surface wind speed RMSE scores. For the partial sunlight episode the MRR configuration was one of two physics suites with the lowest surface temperature RMSE scores, and was among the lowest for the near-total-darkness episode. Of all the physics suites, the MRR package had the lowest warm bias during the most extremely cold periods, both when compared to the surface METAR stations and the Fairbanks sounding. The reason is not known for sure but is probably due to some combination of the effects of its snow model and top-level 'skin' layer. Since the extremely cold conditions are those with the potential for the highest PM2.5 concentrations, we took this as an additional reason to recommend the MRR physics suite for use by EPA.

However, there were periods in each episode, generally when the temperature was steadily decreasing in advance of an extremely cold period, during which all the physics configurations would tend to have a cold bias. During these periods the MRR configuration would still have colder temperatures than the other physics suites, and thus have worse magnitude temperature biases and RMSE scores. The relatively poorer performance of the MRR suite during a such a period accounts for the relatively poorer surface temperature statistics of the MRR suite compared to the MNR suite for the entire near-total-darkness episode. The reason for this behavior is not definitely known, but it is thought to be related to the interaction of radiation with the ice condensate that tends to occur during these periods. Also, the temperature biases of the MRR physics suite during the extremely cold period near the end of the partial sunlight episode were not quite as improved during daylight hours as during nighttime hours as compared to the other physics suites. Therefore, while overall we recommended the MRR configuration to EPA for these episodes, we also strongly recommended that the final fine-scale atmospheric data analysis (i.e., from Grid 3) to be provided to EPA should come from an additional simulation in which FDDA is performed directly on Grid 3, in order to reduce some of this error.

Use of obs nudging for temperature and humidity (and not surface wind) on Grid 3 produced large improvements in the mass fields as expected, and also improvements in the wind fields above the surface.   Results were very encouraging and suggested that a smaller (larger) time window should be used for the surface (above-surface) data assimilation. This capability present in the Penn State MM5 FDDA system has been added to the new-release version of WRF.

In addition to this final report, deliverables to the EPA will include the full three-dimensional output at relatively fine temporal resolution (every 1 hour for Grid 1; every 12 minutes for Grids 2 and 3) for the final Grid 3 nudging simulation as well as all the baseline and physics sensitivity simulations. Model namelists, initialization files, and modifications to the model source code will also be provided.

The development and refinement of WRF FDDA capabilities and supporting software, including the surface analysis nudging, observation nudging and the OBSGRID objective analysis and obs-nudging pre-processing code, occurred concurrently with this project.   This separate development effort led by PI Dave Stauffer and funded by the Defense Threat Reduction Agency (DTRA) allowed us rapid access to the most recent and robust versions of the WRF FDDA code, which greatly benefited this project.

The results of the default FDDA procedures not performing well here in this high vertical resolution modeling study of stable boundary layer environments motivated an additional FDDA code development effort to make the vertical influence functions of surface observations within the FDDA be a function of stability regime type, as well as to provide the user with greater flexibility in specifying the vertical influence functions. These modifications were not finalized in time to be used for this project but are scheduled to appear in the next official release of the WRF model.

An extended abstract and oral presentation were made at the 13th Conference on Mesoscale Processes (Gaudet et al. 2009), and a manuscript based on the project is in preparation.

6.2 Limitations of the Study and Recommendations for Future Work

Sensitivity to the microphysics parameterization was not performed here, but may be important to investigate further. In particular, results from this study suggested that both the occurrence of large negative RUC temperature biases and large differences between the RRTM and RRTMG longwave radiation schemes tended to occur when low-level ice condensate was present.   Therefore, the microphysics / radiation interaction should probably be investigated further.

A fourth grid with 0.44-km horizontal grid spacing centered over Fairbanks was set up and initialized with topography, but it was not used in the sensitivity experiments here.   Although this is finer horizontal resolution than the resolution requested by EPA, some of Penn State's past studies of SBLs (Stauffer et al. 2009) have suggested that the weak wind flows in these conditions may be sensitive to topographic features on these smaller scales, and it might be important to know if finer resolution is also required to resolve the topographic flows of the Fairbanks region.

The latest version of the WRF FDDA code has been designed to have more flexibility in how the temporal and spatial weighting functions are specified. Future simulations that use these new WRF FDDA options that were not yet available for this study should produce a further reduction of model error.

The availability of more meteorological observations from the immediate Fairbanks North Star Borough region, and in particular observations immediately above the surface, would allow one to make a more rigorous assessment of the accuracy of the different physics schemes (in particular, the PBL parameterizations).

More testing and analysis of the model physical parameterizations should be performed to determine the cause of the strong model biases often observed in the simulations, such as the generally persistent warm bias, and the cold RUC land surface model bias during falling temperature conditions.

7. REFERENCES

Benjamin, S.O., and N.L. Seaman, 1985: A simple scheme for objective analysis in curved flow. *Mon. Wea. Rev.*, **113**, 1184-1198.

Benson, C.S., 1970: Ice fog: Low temperature air pollution. Research Report 121. U.S. Army Corps of Engineers, Cold Regions Research and Engineering Laboratory, Hanover, NH, 118 pp.

Bromwich, D.H., J.J. Cassano, T. Klein, G. Heinemann, K.M. Hines, K. Steffen, and J.E. Box, 2001: Mesoscale modeling of katabatic winds over Greenland with the Polar MM5. *Mon. Wea. Rev.*, **129**, 2290-2309.

Chen, F., and J. Dudhia, 2001: Coupling an advanced land-surface/hydrology model with the Penn State/NCAR MM5 modeling system. Part I: Model implementation and sensitivity. *Mon. Wea. Rev.*, **129**, 569-585.

Deng, A., D. Stauffer, B. Gaudet, J. Dudhia, J. Hacker, C. Bruyere, W. Wu, F. Vandenberghe, Y. Liu, and A. Bourgeois, 2009: Update on WRF-ARW end-to-end multi-scale FDDA system. *10th Annual WRF Users' Workshop*, 23 Jun 2009, Boulder, CO.

Dudhia, J., 1989: Numerical study of convection observed during winter monsoon experiment using a mesoscale two-dimensional model. *J. Atmos. Sci.*, **46**, 3077-3107.

Galperin, B., S. Sukoriansky, and P.S. Anderson, 2007: On the critical Richardson number in stably stratified turbulence. *Atmos. Sci. Let.*, **8**, 65-67.

Gaudet, B., D. Stauffer, N. Seaman, A. Deng, K. Schere, R. Gilliam, J. Pleim, and R. Elleman, 2009: Modeling extremely cold stable boundary layers over interior Alaska using a WRF FDDA system.

13th *Conference on Mesoscale Processes*, 17-20 Aug, Salt Lake City, UT, American Meteorological Society.

Girard, E., and J.-P. Blanchet, 2001: Microphysical parameterization of Arctic diamond dust, ice fog, and thin stratus for climate models. *J. Atmos. Sci.*, **58**, 1181-1198.

Hanna, S.R., 1983: Lateral turbulence intensity and plume meandering during stable conditions. *J. Appl. Meteor.*, **22**, 1424-1430.

Hines, K.M., and D.H. Bromwich, 2008: Development and testing of polar Weather Research and Forecasting (WRF) model. Part I: Greenland ice sheet meteorology. *Mon. Wea. Rev.*, **136**, 1971-1989.

Janjić, Z.I., 2002: Nonsingular implementation of the Mellor-Yamada Level 2.5 Scheme in the NCEP Meso model. NCEP Office Note 437, 61 pp.

Mahrt, L., 2009: Characteristics of submeso winds in the stable boundary layer. *Boundary-Layer Meteorology*, **130**, 1-14.

Mlawer, E.J., S.J. Taubman, P.D. Brown, M.J. Iacono, and S.A. Clough, 1997: Radiative transfer for inhomogeneous atmosphere: RRTM, a validated correlated-k model for the longwave. *J. Geophys. Res.*, **102**, 16663-16682.

Mölders, N. and G. Kramm, 2010: A case study on wintertime inversions in interior Alaska with WRF. *Atmos. Res.*, **95**, 314-332.

Morrison, H., J.A. Curry, and V.I. Khvorostyanov, 2005: A new double-moment microphysics parameterization for application in cloud and climate models. Part I: Description. *J. Atmos. Sci.*, **62**, 1665-1677.

Seaman, N.L., and S.A. Michelson, 2000: Mesoscale meteorological structure of a high-ozone episode during the 1995 NARSTO-Northeast study. *J. Appl. Meteor.*, **39**, 384-398.

Seaman, N.L., B. Gaudet, A. Deng, S. Richardson, D.R. Stauffer, J.C. Wyngaard, and L. Mahrt, 2008: Evaluation of meander-like wind variance in high-resolution WRF model simulations of the stable nocturnal boundary layer. *10*[th] *Conference on Atmospheric Chemistry*, 21-24 Jan, New Orleans, LA, American Meteorological Society.

Serreze, M.C., J.D. Kahl, and R.C. Schnell, 1992: Low-level temperature inversions of the Eurasian Arctic and comparison with Soviet drifting station data. *J. Climate*, **5**, 615-629.

Skamarock, W.C., J.B. Klemp, J. Dudhia, D.O. Gill, M. Barker, M.G. Duda, X.-Y. Huang, W. Wang, and J.G. Powers, 2008: A description of the Advanced Research WRF version 3. NCAR Technical Note NCAR/TN475+STR.

Smirnova, T.G., J.M. Brown, and D. Kim, 2000: Parameterization of cold-season processes in the MAPS land-surface scheme. *J. Geophys. Res.*, **105**, 4077-4086.

Stauffer, D.R., and N.L. Seaman, 1994: Multiscale four-dimensional data assimilation. *J. Appl. Meteor.*, **33**, 416-434.

Stauffer, D.R., N.L. Seaman, and F.S. Binkowski, 1991: Use of four-dimensional data assimilation in a limited-area mesoscale model. Part II: Effects of data assimilation with the planetary boundary layer. *Mon. Wea. Rev.*, **119**, 734-754.

Stauffer, D.R., B.J. Gaudet, N.L. Seaman, J.C. Wyngaard, L. Mahrt and S. Richardson, 2009: Sub-kilometer numerical predictions in the nocturnal stable boundary layer. *23rd Conference on Weather Analysis and Forecasting/19th Conference on Numerical Weather Prediction*, 1-5 Jun, Omaha, NE, American Meteorological Society.

Sukoriansky, S. B. Galperin, and V. Perov, 2005: Application of a new spectral theory of stably stratified turbulence to atmospheric boundary layers over sea ice. *Boundary-Layer Meteorology*, **117**, 231-257.

Vickers, D., and L. Mahrt, 2004: Evaluating formulations of stable boundary layer height. *J. Appl. Meteor.*, **43**, 1736-1749.

Wyngaard, J.C., 2004: Toward numerical modeling in the 'Terra Incognita'. *J. Atmos. Sci.*, **61**, 1816-1826.

APPENDIX – SUMMARY OF TASKS

Ten tasks were included in the Statement of Work (SOW) for this project. An overview of the tasks and a summary of the work completed on each of these tasks are provided below:

- Task 1 – Participate in kick-off teleconference in accordance with the SOW.

This took place on 11 Sep 2008. The EPA was provided with the specifications of the nested grid configuration that we proposed in the SOW, and we received in turn particular information about the period and region of study from the EPA.

- Task 2 – Prepare workplan and QA/QC plan in accordance with the SOW.

This was submitted to the EPA during Nov. 2008, along with an updated timetable of deliverables provided during the next monthly teleconference. Included was a description of our proposed simulation plan, choice of baseline physics and grid configuration, and method of simulation.

- Task 3 – Participate in monthly project teleconferences.

We held hour-long teleconferences with the project manager and other scientists at Research Triangle Park and EPA Region 10 (which includes Alaska in its jurisdiction) near the beginning of every month between the kick-off meeting and Jan. 2010. These teleconferences were indispensible for coordinating the needs of EPA with our capabilities and adapting to unforeseen developments as they arose.

- Task 4 – Prepare brief monthly progress reports.

These reports provided to the EPA at the end of every month from Oct. 2008 – Dec. 2009, contained in total most of the information and task completion history found in this report.

- Task 5 – Set up meteorological model and conduct initial baseline testing.

After some minor modifications were made to the proposed model grid configuration to maximize the utility of available data, the final specifications of Grids 1, 2, and 3 were confirmed with the EPA in Feb. 2009; more precise coordination of these grids with a parallel emissions modeling project were completed in May 2009. The data assimilation procedures required for the multiscale multigrid procedure to be used for the project were still being developed for the WRF meteorological model, led by PI Dave Stauffer also working on this contract, which helped expedite the testing and validation of these procedures. Furthermore, the testing results had to be confirmed with the version 3.1 of WRF used for most of this study, released in Apr 2009.   By Jun 2009 we determined that the model components were ready to begin physics sensitivity testing.

- Task 6 – Develop and/or adapt one or more stable boundary layer and land-surface models in accordance with the SOW.

For the choice of stable boundary model in the WRF baseline physics package, we used the Mellor-Yamada-Janjić (MYJ) parameterization that we had used for our previous studies of the stable boundary layer in Alaska, with a few modifications. For the land surface model, however, we decided that we should make use of the Noah model available in version 3.1 of WRF, since it included a number of adaptations to snow-covered terrain that would be critical in this study. Using the particular Noah adaptations in version 3.1 of WRF was one reason for using that model when it became available. After we confirmed that using the Noah land surface model initialized with Global Forecast System (GFS) model data produced reasonable results, we discovered that the default WRF data assimilation procedure needed to be modified to interact properly with the stable boundary layers generated by the high-resolution model. By Jul 2009 we had decided on the baseline physics package to be used for the main simulations.

- Task 7 – Conduct up to five sensitivity tests for the selected modeling periods and evaluate results in accordance with the SOW.

Two twenty-day episodes from the 2007-2008 winter season, both with periods of extremely cold temperatures and high PM2.5 concentrations, were selected for evaluation of model performance: one in near total darkness (14 Dec 2007 – 03 Jan 2008), and the other in partial sunlight (23 Jan 2008 – 12 Feb 2008). In addition to the baseline physics configuration that included the MYJ planetary boundary layer (PBL) scheme, the Noah land surface model, and the RRTM longwave / Dudhia shortwave radiation package, three other physics sensitivity tests were performed for the entirety of each twenty-day episode, which involved using the RRTMG radiation package (longwave and shortwave), the Quasi-Normal Scale Elimination (QNSE) PBL scheme, and the Rapid Update Cycle (RUC) land surface model. After some discussion, the specific combinations used, in addition to the baseline, were MYJ / Noah / RRTMG, QNSE/ Noah / RRTMG, and MYJ / RUC / RRTMG. After statistical comparison with available observations, there was no clearly superior model physics combination; however, the MYJ / RUC / RRTMG option seemed to do the best job at reproducing the extremely cold temperatures characteristic of the high exceedance episodes. However, all model configurations tended to have substantial warm surface temperature biases in these conditions on the innermost 1.3-km Grid 3 when no data assimilation was performed on it. (Data assimilation was performed on the outer Grids 1 and 2 for the physics sensitivity experiments to improve the lateral boundary conditions on Grid 3.) In Jan 2009 it was decided that the MYJ / RUC / RRTMG combination was to be recommended, but that final dynamic analyses using this physics package along with Grid 3 data assimilation should be performed for each episode in order to reduce the atmospheric model errors and biases before they are used in air transport and chemistry models.

- Task 8 – Participate in 1.5-day meeting with Project Officer and scientific staff at EPA/RTP in accordance with the SOW.

This meeting occurred 19-20 Nov 2009 at Research Triangle Park (RTP), NC, between one of the co-PI's (Brian Gaudet) and the Project Officer and other scientific staff at RTP. During this meeting scientific discussion of the results occurred, and a preliminary agreement that the MYJ/RUC/RRTMG combination was the most promising was reached. The main results of the project to date were presented, and plans for bringing the project to completion were made.

- Task 9 – Prepare final report and electronic data and computer code files in accordance with the SOW.

- Task 10 – Revise draft final report and data files.

This report serves to help complete Tasks 9 and 10. A pair of 2-Terabyte external hard drives were obtained from EPA for use for transferring the data, whose cumulative size is approximately 600 Gigabytes per episode simulation. The files to be transferred consist of a full three-dimensional set of model output files, generated every hour for the 12-km Grid 1, and every 12 minutes for the 4-km Grid 2 and 1.3-km Grid 3. The output for each episode from the final dynamic initialization (i.e., with data assimilation on Grid 3) using the best choice physics package will be transferred first; later, the output from the baseline and physics sensitivity studies without Grid 3 data assimilation will be transferred to EPA. In addition, the namelist specifications for each simulation, the WRF version 3.1 code as modified for the project, and the initial, boundary condition, and four-dimensional data assimilation (FDDA) files required for each WRF simulation will be included.