Draft Final Report

Annual 2002 MM5 Meteorological Modeling to Support Regional Haze Modeling of the Western United States

Prepared for

The Western Regional Air Partnership (WRAP)
1515 Cleveland Place, Suite 200
Denver, CO 80202

Prepared by

Sue Kemball-Cook
Yiqin Jia
Chris Emery
Ralph Morris
ENVIRON International Corporation
101 Rowland Way, Suite 220
Novato, CA 94945-5010

Zion Wang
Gail Tonnesen
University of California at Riverside
College of Engineering
Center for Environmental Research and Technology
Riverside, CA 92507

March 2005

TABLE OF CONTENTS

Page

1.0 INTRODUCTION 1-1

Background 1-1
Overview of WRAP MM5 Modeling 1-1

2.0 MM5 SETUP FOR 36 KM CENRAP, VISTAS, and WRAP RUNS 2-1

MM5 Configuration 2-1
Four Dimensional Data Assimilation (FDDA) 2-7

3.0 EVALUATION METHODOLOGY FOR THE 2002 ANNUAL MM5 RUNS 3-1

Graphical Evaluation 3-1
Statistical Evaluation 3-2
The METSTAT Analysis Package 3-5
Statistical Performance Benchmarks 3-6
Surface Statistical Analyses for WRAP 3-6
Precipitation Evaluation 3-8

4.0 SURFACE EVALUATION OF THE 2002 CENRAP, VISTAS, and WRAP RUNS 4-1

Subdomain Surface Field METSTAT Evaluation 4-1
Evaluation of Precipitation in the 36 km 2002 CENRAP, VISTAS, and WRAP Runs 4-17
Upper-Air Meteorological Evaluation 4-26

5.0 EVALUATION OF THE 2002 12 KM WRAP MM5 SIMULATION 5-1

Evaluation of Precipitation in the 2002 12 km WRAP Run 5-6

6.0 CONCLUSIONS 6-1

REFERENCES R-1

TABLES

Table 1-1. MM5 Configuration from original (WRAP_0) 2002 MM5 run 1-3
Table 1-2. MM5 Configuration for the final WRAP 2002 36 km MM5 run. 1-3
Table 2-1. Physics options common to all four MM5 runs. 2-3
Table 2-2. Physics/FDDA options which differed among the four MM5 runs. 2-4
Table 2-3. Final MM5 configuration for the 12-km grid 2-6
Table 3-1. Statistical benchmarks for evaluating meteorological model performance. 3-6
Table 4-1. METSTAT subdomain abbreviations. 4-2

FIGURES

Figure 2-1. Spatial coverage of the RPO National Grid with 36-km grid point spacing. 2-2
Figure 2-2. Coverage of the WRAP western regional domain with 12-km grid point spacing. 2-3
Figure 3-1. METSTAT subdomains for the WRAP 2002 MM5 performance evaluation..... 3-7
Figure 4-1. 36 km surface wind comparison soccerplot for January over western U.S. for VISTAS, CENRAP, WRAP_0 (old WRAP) and final WRAP (WRAP) 2002 36 km simulation..... 4-3
Figure 4-2. 36 km surface wind comparison soccerplot for January over central and eastern U.S. for VISTAS, CENRAP, WRAP_0 (old WRAP) and final WRAP (WRAP) 2002 36 km simulation..... 4-3
Figure 4-3. WRAP_0 run surface wind speed and bias for January over desertSW subdomain 4-4
Figure 4-4. Final WRAP run surface wind speed and bias for January over desertSW subdomain. 4-4
Figure 4-5. WRAP_0 run surface wind direction and bias for January over desertSW subdomain. 4-5
Figure 4-6. Final WRAP run surface wind direction and bias for January over desertSW subdomain 4-5
Figure 4-7. 36 km temperature comparison soccerplot for January over western U.S. for VISTAS, CENRAP, WRAP_0 (old WRAP) and final WRAP (WRAP) 2002 36 km simulation..... 4-6
Figure 4-8. 36 km temperature comparison for January over central and eastern U.S. for VISTAS, CENRAP, WRAP_0 (old WRAP) and final WRAP (WRAP) 2002 36 km simulation..... 4-6
Figure 4-9. WRAP_0 run January temperature time series for desertSW subdomain. 4-7
Figure 4-10. Final WRAP run January temperature time series for desertSW subdomain 4-7
Figure 4-11. 36 km humidity comparison soccerplot for January over western U.S. for VISTAS, CENRAP, WRAP_0 (old WRAP) and final WRAP (WRAP) 2002 36 km simulation..... 4-8
Figure 4-12. 36 km humidity comparison soccerplot for January over central and eastern U.S. for VISTAS, CENRAP, WRAP_0 (old WRAP) and final WRAP (WRAP) 2002 36 km simulation 4-8
Figure 4-13. 36 km wind comparison soccerplot for July over western U.S. for VISTAS, CENRAP, WRAP_0 (old WRAP) and final WRAP (WRAP) 2002 36 km simulation..... 4-11
Figure 4-14. 36 km wind comparison soccerplot for July over central and eastern U.S. for VISTAS, CENRAP, WRAP_0 (old WRAP) and final WRAP (WRAP) 2002 36 km simulation..... 4-11
Figure 4-15. 36 km temperature comparison soccerplot for July over western U.S. for VISTAS, CENRAP, WRAP_0 (old WRAP) and final WRAP (WRAP) 2002 36 km simulation..... 4-12
Figure 4-16. 36 km temperature comparison soccerplot for July over central and eastern U.S for VISTAS, CENRAP, WRAP_0 (old WRAP) and final WRAP (WRAP) 2002 36 km simulation.. 4-12

Figure 4-17. WRAP_0 run July temperature time series for SW subdomain. 4-13
Figure 4-18. Final WRAP run July temperature time series for SW subdomain. 4-13
Figure 4-19. WRAP_0 run July temperature time series for North subdomain. 4-14
Figure 4-20. Final WRAP run July temperature time series for North subdomain. 4-14
Figure 4-21. 36 km humidity comparison soccerplot for July over western U.S. 4-15
Figure 4-22. 36 km humidity comparison soccerplot for July over eastern and central U.S. 4-15
Figure 4-23. January observed (top left) and modeled precipitation for CENRAP (top right), VISTAS (middle left), final WRAP (middle right) and WRAP_0 (bottom right) 2002 36 km simulations and differences between WRAP-0 and final WRAP (bottom left) 4-19
Figure 4-24. Partitioning of January total rainfall into its convective and stratiform components for the WRAP_0 (top) and final WRAP (bottom) 36 km MM5 runs. 4-20
Figure 4-25. March observed (top left) and modeled precipitation for CENRAP (top right), VISTAS (middle left), final WRAP (middle right) and WRAP_0 (bottom right) 2002 36 km simulations and differences between WRAP-0 and final WRAP (bottom left) 4-21
Figure 4-26. July observed (top left) and modeled precipitation for CENRAP (top right), VISTAS (middle left), final WRAP (middle right) and WRAP_0 (bottom right) 2002 36 km simulations and differences between WRAP-0 and final WRAP (bottom left) 4-22
Figure 4-27. Partitioning of July total rainfall into its convective and stratiform components for the WRAP_0 (top) and final WRAP (bottom) 36 km MM5 runs 4-23
Figure 4-28. October observed (top left) and modeled precipitation for CENRAP (top right), VISTAS (middle left), final WRAP (middle right) and WRAP_0 (bottom right) 2002 36 km simulations and differences between WRAP-0 and final WRAP (bottom left). 4-24
Figure 4-29. December observed (top left) and modeled precipitation for CENRAP (top right), VISTAS (middle left), final WRAP (middle right) and WRAP_0 (bottom right) 2002 36 km simulations and differences between WRAP-0 and final WRAP (bottom left) 4-25
Figure 4-30. January 7, 2002, 12Z Sounding for Midland, TX. Upper left hand panel: CENRAP, Upper right hand panel: VISTAS, Lower left hand panel: WRAP_0, Lower right hand panel: Final WRAP 4-29
Figure 4-31. July 2, 2002, 12Z Sounding for Spokane, WA. Upper left hand panel: CENRAP, Upper right hand panel: VISTAS, Lower left hand panel: WRAP_0, Lower right hand panel: Final WRAP. 4-30
Figure 4-32. July 16, 2002, 0Z Sounding for Midland, TX. Upper left hand panel: CENRAP, Upper right hand panel: VISTAS, Lower left hand panel: WRAP_0, Lower right hand panel: Final WRAP. 4-31

Figure 4-33. January 4, 2002, 12Z Sounding for Flagstaff, AZ. Upper left hand panel: CENRAP, Upper right hand panel: VISTAS, Lower left hand panel: WRAP_0, Lower right hand panel: Final WRAP. 4-32
Figure 4-34. July 4,6, 2002, 0Z Soundings for Flagstaff, AZ. Upper left hand panel: WRAP_0 July 4, Upper right hand panel: Final WRAP July 4, Lower left hand panel: WRAP_0 July 6, Lower right hand panel: Final WRAP Jul 6. 4-33
Figure 5-1. 36km/12 km surface temperature comparison soccerplot for January. 5-3
Figure 5-2. 36km/12 km surface wind comparison soccerplot for January 5-3
Figure 5-3. 36km/12 km surface humidity comparison soccerplot for January 5-4
Figure 5-4. 36km/12 km surface temperature comparison soccerplot for July 5-4
Figure 5-5. 36km/12 km surface wind comparison soccerplot for July 5-5
Figure 5-6. 36km/12 km humidity wind comparison soccerplot for July 5-5
Figure 5-7A-J. Annual cycle in 12 km MM5 precipitation 5-7

1.0 INTRODUCTION

BACKGROUND

In 1999, the U.S. Environmental Protection Agency (EPA) announced a major effort to improve visibility in national parks and wilderness areas. Its Regional Haze Rule (RHR) calls for state and federal agencies to work together to improve visibility in 156 Federally mandated Class I areas that includes national parks, wilderness areas and wildlife refuges (referred to as "Class I" areas). The objective of the RHR is to achieve natural visibility conditions in these federally protected lands by the year 2064. The rule requires the states to develop and implement air quality protection plans to reduce the pollution that causes visibility impairment. The Western Regional Air Partnership (WRAP) is a consortium of federal, state, and tribal agencies charged with implementing regional planning processes to improve visibility in western Class I areas. To meet this goal, WRAP is providing the necessary technical and policy tools to help states and tribes implement the RHR.

The WRAP has formed the Regional Modeling Center (RMC), consisting of the University of California at Riverside (UCR), ENVIRON International Corporation and the University of North Carolina, Carolina Environmental Program (UNC/CEP). The RMC performed the modeling and analysis necessary to develop the Section 309 State and Tribal Implementation Plans (SIPs/TIPs) and are performing the analysis for the Section 308 SIPs/TIPs. This project is focused on seasonal-annual regional visibility modeling using a deterministic photochemical grid model capable of simulating the formation, transport, and fate of tropospheric particulate matter (PM) and photochemical oxidants on time scales of one hour. The RMC is utilizing EPA's Models-3 Community Multiscale Air Quality (CMAQ) modeling system and the Comprehensive Air-quality Model with extensions (CAMx) for this purpose. The models will be applied for the entire year of 2002 on two grids: a continental-scale domain with 36-km grid spacing, and a western U.S. regional-scale domain with 12-km grid spacing covering the western U.S. Class I areas. The Fifth Generation Mesoscale Model (MM5), developed and maintained by the Pennsylvania State University and National Center for Atmospheric Research (PSU/NCAR), is being used to supply hourly meteorological fields on these two grids. The Sparse Matrix Operating Kernel Emissions (SMOKE) system is being used to develop the anthropogenic and biogenic gas and PM emission rate estimates for CMAQ and CAMx.

OVERVIEW OF WRAP MM5 MODELING

Initial MM5 Modeling

MM5 simulations were carried out by the WRAP Regional Modeling Center (RMC) for the entirety of 2002 to support visibility modeling for the Section 308 SIP/TIP that are due in December 2007. During the fall of 2003, the RMC made an initial MM5 run for the year 2002 on the single unified National RPO 36-km grid. An MM5 Modeling Protocol was prepared (ENVIRON and UCR, 2004), which describes in detail the MM5 model and the setup and evaluation methods used in the 2002 modeling effort. The model performance evaluation for an initial 2002 MM5 run on a single continental-scale 36-km grid and results from preliminary MM5 sensitivity tests, have been discussed by Morris et al. (2004a), Kemball-Cook et al. (2004a and 2004b), and Emery et al. (2004).

The model configuration for this initial run was based upon a prior 2002 MM5 application undertaken by the Iowa Department of Natural Resources (IDNR; Johnson, 2003 Personal Communication), which was in turn set according to the optimal MM5 physics options that resulted from an in-depth sensitivity project carried out by IDNR and the Lake Michigan Air Directors Consortium (LADCo). While the IDNR simulations used MM5 version 3.5, the applications carried out for WRAP used the latest version of the model (v3.6.1) available at the end of 2003. Additional modifications to the physics configuration and application methodology were made for the WRAP simulation based on the latest information from EPA, IDNR, LADCo, and others. First, the RMC opted to use the Reisner II mixed-phase cloud microphysics package according to suggestions from EPA/ORD. Second, the INTERPPX option, which allows for continuous soil moisture initialization from one run segment to the next, was not used based on poor performance reported by IDNR and LADCo.

The initial 2002 36-km WRAP simulation (referred to hereafter as WRAP_0) results showed that MM5 performs better in the Central and Eastern U.S. than in the West, and performs generally better in winter than in summer (Morris et al., 2004a; Kemball-Cook et al., 2004a). In the western U.S., the amplitude of diurnal temperature cycle was persistently underestimated during the summer, especially in the southwest. In the desert southwest, the humidity was greatly overestimated during the summer as well, and there was a pronounced cold bias. Some of these problems appeared to be linked to the excessive simulated precipitation generated by MM5 during the summer, especially in the southwest. This can have serious repercussions for CMAQ modeling since too much rain can “wash out” pollutants, while the too cool, humid and cloudy environment may lead to incorrect pollutant chemistry and aerosol thermodynamics. Temperature and humidity problems overshadowed the surface wind performance, which was not particularly good, but was likely affected by smaller-scale topographic influences that are not well represented even at the finer 12 km resolution. Wind performance improved quickly with height above the surface, suggesting that regional transport speeds/directions were reasonably represented.

Preliminary Sensitivity Tests

Initial sensitivity tests were undertaken for a 5-day period in July 2002 for the 36-km National domain in an attempt to find alternative MM5 configurations that would improve the poor summertime performance in the western U.S. Further analyses were conducted to compare the WRAP_0 run’s performance against that of a 36-km MM5 run made for VISTAS, and against operational Eta Data Assimilation System (EDAS) fields used as input to MM5. From these analyses, we found that the use of the Kain-Fritsch II scheme (as used in the VISTAS 2002 MM5 modeling compared to the Kain-Fritsch I scheme used in the initial WRAP modeling) improved, but did not entirely solve, the precipitation over prediction problem in the WRAP_0 run. We also found that removal of soil moisture nudging improved temperature and humidity performance for the short summertime tests. Evaluation of the EDAS fields used in the MM5 four-dimensional-data-assimilation (FDDA) revealed that EDAS did not exhibit the summer time cold wet bias in the southwest. This indicated that the bias was not introduced by the FDDA. A new model configuration (Run 5) was identified from these initial tests for new annual 2002 simulations, and the WRAP MM5 modeling protocol was updated to reflect this, as shown in the Table 1-1.

Table 1-1. MM5 Configuration from original (WRAP_0) 2002 MM5 run.

	LSM	PBL	Cumulus	Microphysics	Analysis FDDA 3D	Analysis FDDA Surface	Obs FDDA	Soil Moisture Nudging
WRAP_0	PX	ACM	KF I	Reisner II	W/T/H	W/T/H	None	Yes
WRAP_5	PX	ACM	KFII	Reisner II	W/T/H		None	No

Although the WRAP_5 configuration did improve humidity, temperature, and precipitation performance over the short July test period (especially where improvement was needed most), the desert southwest continued to exhibit unsatisfactory levels of precipitation and humidity. As stated above, these issues will likely play crucial roles in the air quality model performance. When results of the interim MM5 model run were presented at the May 24-25, 2004 National RPO modeling meeting in Denver, Colorado (Emery et al, 2004), there were still concerns about the 2002 MM5 model performance in the western U.S. Thus, WRAP requested that the Regional Modeling Center (RMC) perform further MM5 sensitivity tests to identify a better performing configuration, including investigating alternative Land Surface Model and Planetary Boundary Layer (LSM/PBL) configurations to Plein-Xiu/Asymmetric Convective Mixing (PX/ACM) schemes used in the initial WRAP MM5 modeling and include the possibility of using different MM5 configurations for different times of year.

Revised Sensitivity Tests

We carried out additional 36-km MM5 test simulations in an attempt to further improve MM5 performance. The latest available version of MM5 (v3.6.2) was used. Since the most severe problems for the initial WRAP 2002 MM5 annual simulation were the wet bias and precipitation overestimation in summer, we first investigated these issues for the 5-day July test period (July 1-5, 2002) on the 36-km grid to maximize computational efficiency. We tested four physics options: (1) the cumulus parameterization; (2) the land surface models (LSM); (3) the planetary boundary layer (PBL) models; and (4) the four-dimensional data assimilation (FDDA).

We then undertook a series of tests to determine the optimal MM5 configuration for the nested 12-km grid of the WRAP modeling domain. In these tests, we varied only the physics options that we expected to ameliorate the biases in the WRAP region seen in the 36-km simulation. The over prediction of summer rainfall and associated surface humidity bias suggested that the cumulus parameterization might need to be changed. Another possible reason for the wet bias was the soil moisture initialization. Therefore, we focused on sensitivity tests of the soil moisture specification and the cumulus scheme. The results of these sensitivity tests are discussed in Kemball-Cook et al. (2004b). The more optimal model configuration for the 36-km application is given in Table 1-2.

Table 1-2. MM5 Configuration for the final WRAP 2002 36 km MM5 run.

LSM	PBL	Cumulus	Microphysics	Analysis FDDA 3D	Analysis FDDA Surface	Obs FDDA
Pleim-Xiu	ACM	Betts-Miller	Reisner II	W/T/H	W/T/H	Wind

MM5 runs for the test period of July 1-5, 2002 made with these physics options selected showed:

- A dramatic reduction in the summertime cold, wet bias in the desert southwest;
- Surface temperature and humidity performance within benchmarks for all WRAP regions except desert southwest for temperature;
- More accurate representation of the diurnal temperature cycle in the desert southwest;
- Improvements in the index of agreement for temperature, humidity and wind speed for all WRAP subdomains;
- A more realistic precipitation pattern over the western U.S.;
- Better model performance in the eastern U.S.; and
- Runs made with these physics options generally represented an improvement in performance throughout the year, making it unnecessary to select different physics schemes for different seasons.

In summary, the new configuration of MM5 represented a significant improvement in model performance relative to the WRAP_0 and WRAP_5 2002 36-km MM5 runs for the 5 day test period.

We then conducted a series of tests to determine the optimal MM5 configuration for the nested 12-km grid of the WRAP modeling domain. In these tests, we again focused on sensitivity tests of the soil moisture specification and the cumulus scheme.

The main results from the 12-km runs over the test period of July 1-5, 2002 are as follows:

- Use of soil moisture nudging degraded temperature and humidity model performance.

- Initializing the soil moisture with the 36-km domain MM5 output degraded temperature and humidity model performance.

- Better temperature and humidity performance was obtained by initializing the soil model with the EDAS analysis fields.

Final Model Configurations

Based upon our evaluation of surface variable and rainfall performance, we selected a configuration with the Betts-Miller cumulus scheme running on the 36-km grid and no cumulus parameterization on the 12-km grid. This was among the best performers in terms of surface wind, temperature, and humidity, and performed the best for rainfall. We also chose to remove surface analysis nudging of temperature and humidity from both grids, based on guidance from reviewers that nudging to these variables can degrade performance aloft. Other MM5 model options remained as given in Table 1-2. The final 36/12-km configuration was then used for the annual 2002 MM5 simulation to support WRAP's regional haze modeling. The WRAP Modeling Protocol (ENVIRON and UCR 2004) was updated to reflect the new model configuration.

Report Organization

In this document, we evaluate the final continental-scale 36-km grid MM5 run performed by the WRAP RMC during 2004 and compare its performance with the initial WRAP_0 run as well as with two additional 36 km continental-scale annual runs performed by the CENRAP and VISTAS RPOs. First, we present brief descriptions of the MM5 physical configuration used in each MM5 run. Then, we evaluate the performance of each model run in replicating the evolution of observed winds, temperature, humidity, and boundary layer morphology to the extent that resources and data availability allow; this serves as an assessment of the reliability of the final WRAP run's 36-km meteorological fields in adequately characterizing the state of the atmosphere and for serving as boundary conditions for the 12-km regional-scale WRAP domain MM5 run and the meteorological driver for the CMAQ 36-km continental-scale run. Finally, we evaluate the final 12 km WRAP run and discuss its suitability for providing meteorological conditions for the proposed 12 km CMAQ application.

2.0 MM5 SETUP FOR 36 KM CENRAP, VISTAS, and WRAP RUNS

In this section, we define the configurations for the CENRAP, VISTAS, WRAP_0 (initial WRAP) and final WRAP 2002 36 km and 12 km MM5 simulations.

MM5 CONFIGURATION

The CENRAP modeling configuration is described in Johnson (2004), and the VISTAS run is described in Olerud and Sims (2003). The WRAP MM5 modeling system configuration for the 2002 36 km and 12 km annual runs is described in the WRAP RMC MM5 Modeling Protocol (ENVIRON and UCR 2004). Additional modifications to the physics configuration and application methodology were made for the WRAP MM5 simulation based on the results of a sensitivity study aimed at improving the performance problems seen in the first WRAP run (Kemball-Cook et al. 2004).

Modeling Domain

In all four 36 km runs (CENRAP, VISTAS, WRAP_0 and final WRAP), MM5 was configured to run on the standard continental-scale Regional Planning Organization (RPO) National Grid with 36-km grid point spacing (Figure 2-1). The RPO National Grid is defined on a Lambert conformal projection, with true latitudes at 33°N and 45°N, and the central latitude and longitude at 40°N and 97°W, respectively. The grid point spacing is 36 km. The continental expanse of this domain results in a grid of 165 (east-west) by 129 (north-south) dot points, and 164 (east-west) by 128 (north-south) cross points (Figure 2-1). Overall, the domain covers 5904 km by 4608 km. The 12 km WRAP domain is shown in Figure 2-2.

Vertical Grid Structure

The vertical layer structure of the WRAP_0 and final WRAP runs is detailed in the Modeling Protocol. The vertical structure for the CENRAP and VISTAS runs was similar to that of the two WRAP runs. Both CENRAP and VISTAS runs used 34 layers in the vertical, but the exact layer definitions varied slightly.

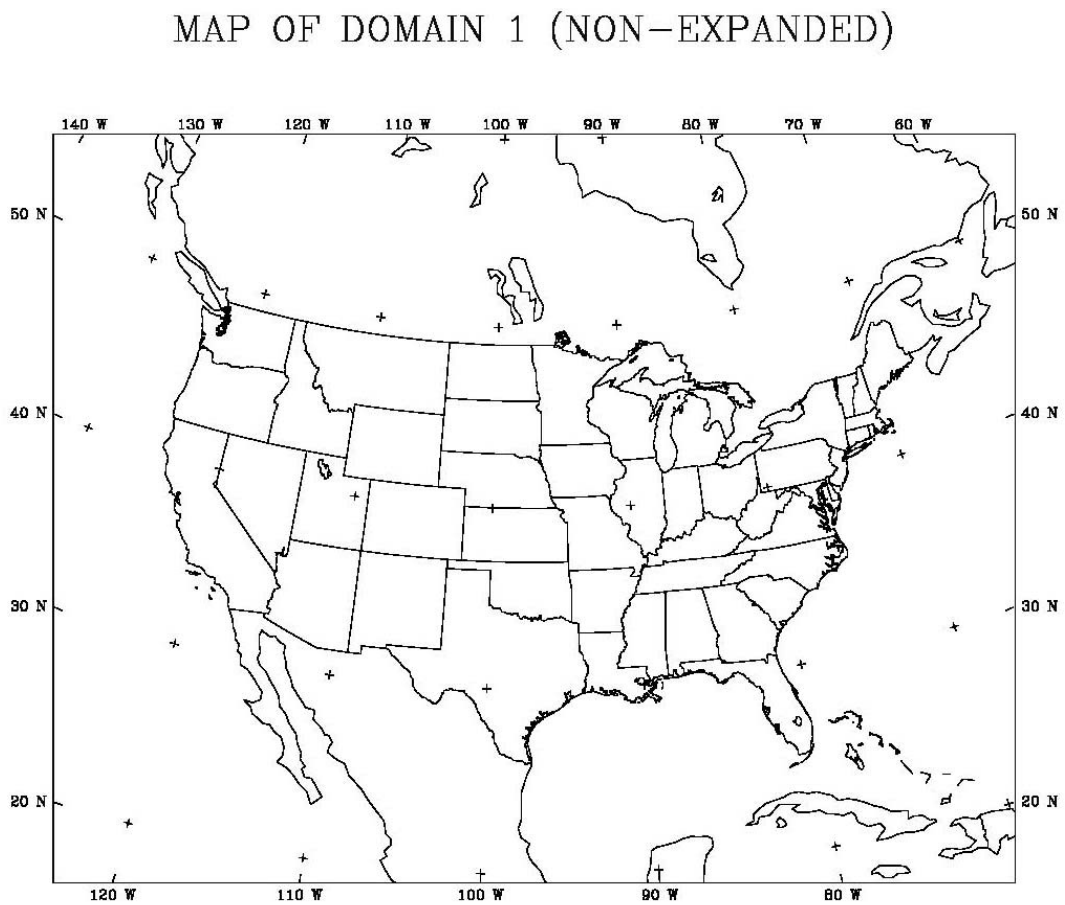

Figure 2-1. Spatial coverage of the RPO National Grid with 36-km grid point spacing.

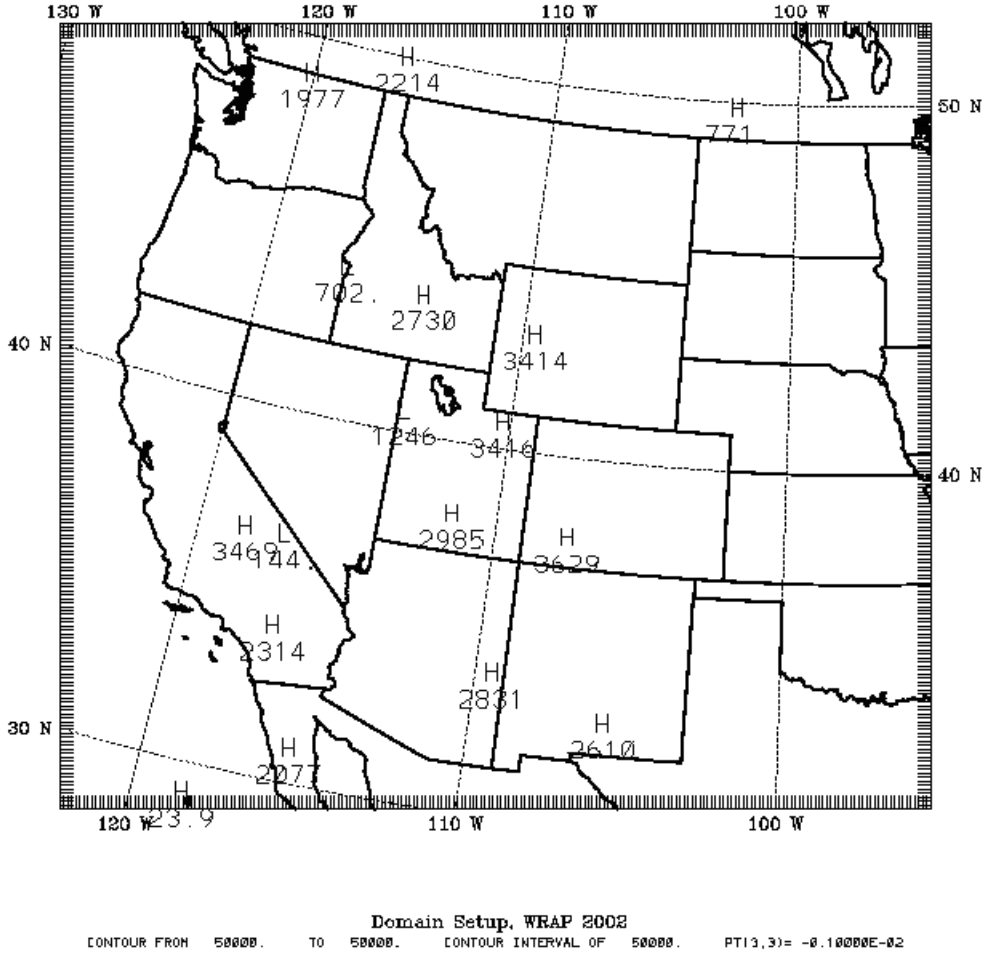

Figure 2-2. Coverage of the WRAP western regional domain with 12-km grid point spacing.

Physics Parameterizations and FDDA

In this section, we describe the major similarities and differences between the three model runs. Table 2-1 shows the physics options common to all four runs, whereas Table 2-2 displays these differences.

Table 2-1. Physics options common to all four MM5 runs.

Land Surface Model	Pleim-Xiu LSM (Used without interppx option in all four runs)
Planetary Boundary Layer	ACM Planetary Boundary Layer
Radiation	RRTM longwave radiation scheme, Dudhia shortwave radiation scheme
Shallow Convection	No shallow convection

Table 2-2. Physics/FDDA options which differed among the four MM5 runs.

	CENRAP	VISTAS	WRAP_0	WRAP_Final
Cumulus convection	Kain-Fritsch II	Kain-Fritsch II	Kain-Fritsch I	Betts-Miller
Moist Physics	Reisner I	Reisner I	Reisner II	Reisner II
Analysis Nudging at the Surface	U/V	U/V	U/V/T/Q	U/V
Analysis Nudging Aloft	U/V/T/Q	U/V/T/Q	U/V/T/Q	U/V/T/Q
Surface Obs Nudging				W

Land Surface Model/Planetary Boundary Layer Model

The Pleim-Xiu (PX) parameterization (Xiu and Pleim 2000) is a predictive/interactive soil temperature and moisture budget model that responds to atmospheric processes that affect the thermodynamics of the surface (e.g., rainfall) while in turn dictating the surface fluxes of momentum, heat and moisture into the boundary layer to further affect atmospheric processes. The PX approach maintains a historical "memory" of the soil conditions over the course of a continuous simulation.

Through selection of the Pleim-Xiu land surface model (LSM) all four runs were required to use the Asymmetric Convective Mixing (ACM) Planetary Boundary Layer (PBL) option (Pleim and Chang 1992), as these two are directly coupled. The INTERPX option, which may be employed in multi-segment runs to re-initialize the soil temperature and moisture fields for the next model run segment according to the conditions at the end of the previous run, was not used in any of the four runs. This was in response to sensitivity studies by the IDNR and VISTAS (Olerud and Sims, 2003), which showed that use of INTERPPX can lead to surface temperature and moisture biases that the model physics cannot overcome if the model is initialized during a period of extreme cold. Instead, the PX scheme was run for each simulation segment with the initial soil moisture interpolated from the input EDAS objective analyses fields.

Radiation / Shallow Convection

All four model runs used the RRTM (Rapid Radiative Transfer Model) package for longwave radiation and the Dudhia shortwave radiation option. Shallow convection was turned off.

Cumulus Convection

The parameterization of cumulus convection was a critical factor in model performance in the 2002 annual application. Most of the summer rainfall in the U.S. is due to convective clouds; stratiform rain produced by mid-latitude cyclones is a smaller fraction of the total rainfall in summer, as the mid-latitude storm track shifts northward of most of the U.S.

The WRAP_0 run used the Kain-Fritsch I cumulus parameterization (Kain and Fritsch 1990). Unlike many cumulus schemes that were designed for use in handling deep tropical convection

in general circulation models (i.e. Arakawa-Schubert, Kuo, Betts-Miller), the KF I and II schemes were designed with mid-latitude convection in mind. The Kain-Fritsch I scheme is a mass flux scheme whose closure assumption is that the convective mass flux is constrained to remove grid-scale instability in a specified time period, which is often taken to be the advective time scale (about an hour). The KF I scheme has a complex cloud model that treats both updrafts and downdrafts and can entrain or detrain throughout the depth of the convecting layer. (The Fritsch-Chappell scheme, on which KF I was based, detrains only at the cloud top). Entrainment and detrainment in KF I are handled through a buoyancy-sorting algorithm, in which parcels of cloudy and environmental are mixed, and their buoyancy evaluated. Positively buoyant parcels are entrained, while negatively buoyant parcels are detrained. The KF I cloud model includes ice phase physics. The Kain-Fritsch II (KF II) scheme extends KF I by adding a treatment of shallow convection. The CENRAP and VISTAS 36 km runs both used the KFII cumulus scheme, whereas the WRAP_0 run used KFI.

The final WRAP 36 km run used the Betts-Miller cumulus scheme (Betts and Miller 1986). The Betts-Miller scheme was originally designed to parameterize the effects of deep tropical convection on grid scale variables in general circulation models with grid sizes that could be hundreds of kilometers across. The scheme is based on observations of tropical convection that showed that radiative and convective processes drive the atmosphere toward a vertical temperature and moisture structure that is close to a moist adiabat passing through the equivalent potential temperature of the subcloud layer. The Betts-Miller scheme was designed to reproduce the observed quasi-equilibrium between large scale forcing and convection without resorting to a complicated cloud model, as cloud models are complex and computationally intensive and have closure parameters which may not be well constrained by observations. Instead, the Betts-Miller scheme relaxes the convecting grid column toward reference profiles of temperature and moisture. Betts-Miller includes separate schemes for shallow (i.e. non-precipitating) and deep (precipitating) convection.

The deep convection scheme computes a reference profile that includes the virtual effects of liquid water loading up to the freezing level and adjusts the column toward it with a specified time constant. The subcloud layer is adjusted toward an evaporatively driven downdraft profile with its own timescale derived from an evaporation efficiency. The temperature and moisture reference profiles are different over land and water, and the reference profiles can be further suited to a particular convective environment through a choice of profile based on a cloud efficiency factor. Precipitation from the deep convective scheme is calculated as the integrated residual water between the large-scale moisture profile and the reference profile.

When it comes to using a cumulus parameterization in MM5 at mid-latitudes at ~36 km resolution, the standard choices are the Kain-Fritsch or Grell schemes (MM5 on-line documentation). Betts-Miller, designed for use in the tropics, has some difficulties in the mid-latitudes, and is a less common choice. For example, Hart (2000) notes:

> "Betts-Miller (which has been used in the ETA and mesoETA models) is an undesirable parameterization for mid-latitude mesoscale convective systems. The scheme is too slow to respond to convective instability, often producing rainfall rates which are an order of magnitude less than observed and several hours too late. Further, the scheme cannot respond to elevated instability above the boundary layer. During such cases, the scheme responds too strongly to the low-level unstable air in the warm sector, which then robs moisture from the areas of elevated instability, typically to the north".

In response to the initial WRAP run's performance difficulties in simulating the summer rainfall over the western U.S., we tested MM5's sensitivity to the choice of cumulus parameterization in the WRAP_0 configuration. The original 2002 MM5 run (WRAP_0) used the Kain-Fritsch I scheme. Results from VISTAS 36 km MM5 modeling (Olerud and Sims 2003) showed improved surface fields relative to the WRAP_0 run using Kain-Fritsch II, so Kain-Fritsch II (KFII) was the first scheme tested. We also made similar runs that used the Grell and Betts-Miller parameterizations. All of these runs were made using the PX/ACM LSM/PBL schemes and W/T/Q analysis nudging at the surface and aloft and no nudging to observations (obs nudging). We found that for July, the Betts-Miller scheme gave the best performance in terms of rainfall and surface temperature, humidity and winds, and selected this scheme for use in the final WRAP run, despite the fact that KF is more traditionally used for mid-latitudes.

As part of the WRAP sensitivity modeling, we ran additional tests to evaluate the effect of different cumulus parameterizations on the 12-km WRAP domain. At the 12 km scale, the problem of cumulus-parameterization is not well-posed, as there is no clear spectral gap between the resolved grid-scale process and the scale of the parameterized process (Arakawa and Chen 1987). Molinari (1993) suggests that parameterization of cumulus convection for grid sizes of 2-20 km cannot be addressed with either the fully explicit method (i.e. no cumulus parameterization) or the hybrid parameterization approach (e.g., KFI, KFII, and Betts-Miller). However, our ultimate goal of performing CMAQ visibility modeling requires us to be pragmatic and configure MM5 in a physically reasonable way that produces the most accurate representation of the 2002 meteorology. Our plan for the selection of the 12 km grid cumulus scheme, therefore, was to conduct several sensitivity tests to determine which cumulus scheme (or no cumulus scheme at all) gives the best performance in terms of rainfall, surface temperature, humidity and wind. The best MM5 performance was obtained using no cumulus scheme at all on the 12-km. We recognize that this result could be serendipitous (i.e., leading to a better answer for the wrong reason), however the alternative (poorer performance) is obviously worse and the overstated precipitation would affect the air quality model. The other physics options for the 12-km grid were identical to those used on the 36-km grid. Table 2-3 shows the selected final MM5 configuration for the 12-km application.

Table 2-3. Final MM5 configuration for the 12-km grid.

LSM	PBL	Cumulus	Microphysics	Analysis FDDA		Obs FDDA
				3D	Surface	
Pleim-Xiu	ACM	None	Reisner II	W/T/H	W	Wind

Explicit Moisture Physics

The VISTAS and CENRAP runs used the Reisner I explicit moisture physics scheme (Reisner 1998). The Reisner I scheme employs a bulk cloud microphysics parameterization to predict cloud water, cloud ice, rain, and snow fields. It treats supercooled water and allows for the time evolution of melting snow. The Reisner II scheme builds on the Reisner I scheme by predicting the graupel mixing ratio field, and by adding ice number concentration prediction equations. Although the Reisner I and II schemes are more expensive than the "simple ice" option, the EPA recommends that a mixed phase ice scheme be employed in MM5 to drive aqueous chemistry and wet scavenging in CMAQ. This is because the simple ice approach treats all condensed water forms as a single liquid variable, which, when passed to CMAQ, overstates the quantity of

liquid cloud and precipitation water available for chemistry and removal. Furthermore, the new versions of CMAQ (v4.4) and MCIP (v2.3) now allow the graupel mixing ratio to be passed from MM5 to CMAQ. Thus, the Reisner II was chosen for the final WRAP 36/12 km runs in order to take advantage of this new capability in future visibility modeling.

FOUR DIMENSIONAL DATA ASSIMILATION (FDDA)

In all four 36 km runs, MM5 was configured to use its FDDA capabilities to nudge the model toward wind, temperature, and moisture analysis fields throughout the 2002 annual simulation. Analysis (or grid) nudging was performed in all model runs at 3-hourly intervals both for the surface wind fields and for the wind, moisture, and temperature fields aloft, excluding the boundary layer. Exclusion of the boundary layer in the FDDA process removed the potential for damping resolved mesoscale forcings in the model that are important to boundary layer development and thus the vertical fluxes of momentum, heat, and moisture between the free atmosphere and the surface. The WRAP_0 run used additional surface analysis nudging of temperature and humidity at the surface, and the final WRAP run used observational nudging of winds (to ds472) surface meteorological observational database at the surface. The final WRAP run did not use surface analysis nudging of temperature and humidity at the surface, based on reviewer comments on the Sensitivity Study Report (Kemball-Cook et al. 2004b). Reviewer comments are available at: http://pah.cert.ucr.edu/aqm/308/mm5_reports.shtml.

3.0 EVALUATION METHODOLOGY FOR THE 2002 ANNUAL MM5 RUNS

The goal of the MM5 evaluation is to determine whether the meteorological fields generated by the four 36 km 2002 MM5 simulations and the 12 km WRAP simulation are sufficiently accurate to properly characterize the transport, chemistry, and removal processes for regional haze modeling in the western U.S. If errors in the meteorological fields are too large, the ability of the air quality model to replicate regional pollutant levels over the entire base year will be severely hampered and the predicted air quality and visibility impacts from future year growth and controls will be highly questionable. To provide a reasonable meteorological characterization to the photochemical/visibility model, MM5 must represent with some fidelity the:

- Large-scale weather patterns (i.e., synoptic patterns depicted in the 850-300 mb height fields), as these are key forcings for mesoscale circulations;
- Mesoscale and regional wind, temperature, PBL height, humidity, and cloud/precipitation patterns;
- Mesoscale circulations such as sea breezes and mountain/drainage circulations;
- Diurnal cycles in PBL depth, temperature, and humidity.

For visibility applications, the moisture and condensate fields are particularly important as they significantly impact PM chemical formation, removal, and light scattering efficiency. In addition, cloud and precipitation fields are a good measure of the integrated performance of the model since these are model-derived quantities and are not nudged to observations. Because of the model's coarse resolution of 36 km/12 km, the present runs cannot be expected to faithfully simulate the high frequency pattern or variability of the convective precipitation, but should reproduce the synoptic precipitation and cloud patterns.

In this study, the basis for the operational performance assessment entailed a comparison of the predicted meteorological fields to available surface and aloft meteorological data that are collected, analyzed, and disseminated by the National Weather Service. The performance evaluation was carried out both graphically and statistically to evaluate model performance for winds, temperatures, humidity, and the placement, intensity, and evolution of key weather phenomena.

GRAPHICAL EVALUATION

Because of the regional nature of the haze problem, and the fact that most pollutant transport of interest occurs above the surface, evaluation of the MM5 upper air meteorological fields assumed particular importance in this study. Analysis of MM5 tropospheric output fields is made more difficult by the spatial and temporal resolution of the upper air sounding network, which is sparse in time and space compared with the surface observing network. Radiosonde measurements are usually taken twice per day by an observing network of approximately 120 stations in the continental U.S. Hourly radar profiler measurements are available from NOAA, but are lacking in spatial coverage. Additional plotting capabilities have been developed by the

IDNR (RAOBPLOT) to compare sounding data at individual sites and times to predicted soundings. This allows for a site-by-site comparison of wind, temperature, and humidity profiles, and also provides the user the capability to diagnose and evaluate the heights, strengths, and depths of key stability regimes (e.g., boundary layer depths, nocturnal inversion heights) from observational soundings and MM5 predictions. The observed soundings were derived from the Forecast Systems Laboratory (FSL)/ National Climatic Data Center (NCDC) Dataset.

STATISTICAL EVALUATION

Johnson (2003) provides a synopsis of the challenges associated with undertaking an objective and meaningful performance evaluation. Climatic variability, complex mesoscale phenomenon, stochastic variability, and scientific limitations contribute to unique performance issues in each meteorological modeling exercise, thereby forcing modelers to take a subjective approach to model performance evaluation. Objective statistics that offer a quantitative model assessment exist, but implementation of the metrics is subjective to a degree. For example, defining the area over which domain-averaged metrics are calculated is a subjective decision, buffered only through guidelines. In general, metrics averaged over large modeling domains are avoided, as error cancellation dilutes relevance. Conversely, splitting the modeling domain into small analysis sub-domains can render sample sizes unrepresentative. The logical approach falls well within the bounds of the extremes, leaving optimum sub-domain definitions open to interpretation.

Within the statistical degrees of freedom available to the meteorological modeler, a subset of standard statistical measures has emerged, as described below. These metrics are calculated on hourly and daily time frames for wind speed, wind direction, temperature, and humidity at the surface. Below we list and describe the various statistical measures that will be employed. While no strict criteria establishing acceptable model performance exist, general guidelines have been developed by Emery et al. (2001) and are described later.

<u>Mean Observation (M_o):</u> calculated from all sites with valid data within a given analysis region and for a given time period (hourly or daily):

$$M_o = \frac{1}{IJ}\sum_{j=1}^{J}\sum_{i=1}^{I} O_j^i$$

where O^i_j is the individual observed quantity at site i and time j, and the summations are over all sites (I) and over time periods (J).

<u>Mean Prediction (M_p):</u> calculated from simulation results that are interpolated to each observation point used to calculate the mean observation (hourly or daily):

$$M_p = \frac{1}{IJ}\sum_{j=1}^{J}\sum_{i=1}^{I} P_j^i$$

where P^i_j is the individual predicted quantity at site i and time j. Note that mean observed and predicted winds are vector-averaged (for east-west component u and north-south component v), from which the mean wind speed and mean resultant direction are derived.

Least Square Regression: performed to fit the prediction set to a linear model that describes the observation set for all sites with valid data within a given analysis region and for a given time period (daily or episode). The y-intercept a and slope b of the resulting straight line fit are calculated to describe the regressed prediction for each observation:

$$\hat{P}_j^i = a + bO_j^i$$

The goal is for a 1:1 slope and a "0" y-intercept (no net bias over the entire range of observations), and a regression coefficient of 1 (a perfect regression). The slope and intercept facilitate the calculation of several error and skill statistics described below.

Bias Error (B): calculated as the mean difference in prediction-observation pairings with valid data within a given analysis region and for a given time period (hourly or daily):

$$B = \frac{1}{IJ}\sum_{j=1}^{J}\sum_{i=1}^{I}\left(P_j^i - O_j^i\right)$$

Gross Error (E): calculated as the mean *absolute* difference in prediction-observation pairings

$$E = \frac{1}{IJ}\sum_{j=1}^{J}\sum_{i=1}^{I}\left|P_j^i - O_j^i\right|$$

with valid data within a given analysis region and for a given time period (hourly or daily):
Note that the bias and gross error for winds are calculated from the predicted-observed residuals in speed and direction (not from vector components u and v). The direction error for a given prediction-observation pairing is limited to range from 0 to ±180°.

Root Mean Square Error (RMSE): calculated as the square root of the mean squared difference in prediction-observation pairings with valid data within a given analysis region and for a given time period (hourly or daily):

$$RMSE = \left[\frac{1}{IJ}\sum_{j=1}^{J}\sum_{i=1}^{I}\left(P_j^i - O_j^i\right)^2\right]^{1/2}$$

The RMSE, as with the gross error, is a good overall measure of model performance. However, since large errors are weighted heavily (due to squaring), large errors in a small subregion may produce a large RMSE even though the errors may be small and quite acceptable elsewhere.

Systematic Root Mean Square Error ($RMSE_S$): calculated as the square root of the mean squared difference in regressed prediction-observation pairings within a given analysis region and for a given time period (hourly or daily):

$$RMSE_S = \left[\frac{1}{IJ}\sum_{j=1}^{J}\sum_{i=1}^{I}\left(\hat{P}_j^i - O_j^i\right)^2\right]^{1/2}$$

where the regressed prediction is estimated for each observation from the least square fit described above. The $RMSE_S$ estimates the model's linear (or systematic) error; hence, the better the regression between predictions and observations, the smaller the systematic error.

Unsystematic Root Mean Square Error ($RMSE_U$): calculated as the square root of the mean squared difference in prediction-regressed prediction pairings within a given analysis region and for a given time period (hourly or daily):

$$RMSE_U = \left[\frac{1}{IJ} \sum_{j=1}^{J} \sum_{i=1}^{I} \left(P_j^i - \hat{P}_j^i \right)^2 \right]^{1/2}$$

The unsystematic difference is a measure of how much of the discrepancy between estimates and observations is due to random processes or influences outside the legitimate range of the model.

A "good" model will provide low values of the RMSE, explaining most of the variation in the observations. The systematic error should approach zero and the unsystematic error should approach RMSE since:

$$RMSE^2 = RMSE_S^2 + RMSE_U^2$$

It is important that RMSE, $RMSE_S$, and $RMSE_U$ are all analyzed. For example, if only RMSE is estimated (and it appears acceptable) it could consist largely of the systematic component. This error might be removed through improvements in the model inputs or use of more appropriate options, thereby reducing the error transferred to the photochemical model. On the other hand, if the RMSE consists largely of the unsystematic component, this indicates that further error reduction may require model refinement (new algorithms, higher resolution grids, etc.), or that the phenomena to be replicated cannot be fully addressed by the model. It also provides error bars that may be used with the inputs in subsequent sensitivity analyses.

Index of Agreement (IOA): calculated following the approach of Willmont (1981). This metric condenses all the differences between model estimates and observations within a given

$$IOA = 1 - \left[\frac{IJ \cdot RMSE^2}{\sum_{j=1}^{J} \sum_{i=1}^{I} \left| P_j^i - M_o \right| + \left| O_j^i - M_o \right|} \right]$$

analysis region and for a given time period (hourly and daily) into one statistical quantity. It is the ratio of the total RMSE to the sum of two differences - between each prediction and the observed mean, and each observation and the observed mean:

Viewed from another perspective, the index of agreement is a measure of the match between the departure of each prediction from the observed mean and the departure of each observation from the observed mean. Thus, the correspondence between predicted and observed values across the domain at a given time may be quantified in a single metric and displayed as a time series. The

index of agreement has a theoretical range of 0 to 1, the latter score suggesting perfect agreement.

Most of statistics used to evaluate meteorological model performance are given in absolute terms (e.g., wind speed error in m/s), rather than in relative terms (percent error) as is commonly shown for air quality assessments. The major reason for this is that a very different significance is associated with a given relative error for different meteorological parameters. For example, a 10% error for wind speed measured at 10 m/s is an absolute error of 1 m/s, a minor error. Yet a 10% error for temperature at 300° K is an absolute error of 30° K, a ridiculously large error. On the other hand, pollutant concentration errors of 10% at 1 ppb or 10 ppm carry similar significance.

THE METSTAT ANALYSIS PACKAGE

ENVIRON has developed a statistical analysis software package to calculate and graphically present the statistics described above. The package is comprised of a single Fortran program (METSTAT) to generate observation-prediction pairings and to calculate the statistics, and a Microsoft Excel macro (METSTAT.XLS) that plots the results.

The Fortran program reads MM5 output prediction files and surface observational data files. The program reads either MM5 observation FDDA input files directly, or observation data in an ASCII format. The program then spatially and temporally pairs MM5 predictions with observations for a user-defined time and space window. Since the surface layer in MM5 is usually rather thick relative to the heights at which the observational data were recorded, the METSTAT program includes a micro-meteorological module that scales mid-layer predicted winds to 10 m heights, and mid-layer predicted temperatures to 2 m heights, using common stability-dependent similarity relationships. The horizontal analysis range can be given for an entire MM5 grid, by an LCP coordinate box, or as a list of specific site identifiers (such as WBAN or AIRS numbers), as labeled on the observational file. This allows for an evaluation at a single site, a subset of specific sites (e.g., those along a coastline that would be difficult to select by defining an LCP box) or over an entire regional domain.

The program then proceeds to calculate the statistics described above for each hour and for each day of the time window. The following parameters are considered:

Wind Speed, Temperature, Humidity:
- Mean Observed
- Mean Predicted
- Bias
- Gross Error
- RMSE
- RMSES
- RMSEU
- IOA

Wind Direction:
- Mean Observed
- Mean Predicted
- Bias
- Gross Error

The RMSE and IOA have not been typically used to quantify error for wind direction, and thus are not calculated by the program.

Separate ASCII files containing the hourly and daily statistics are generated, formatted specifically to facilitate import into the Excel macro. The Excel macro is used to plot the data. The hourly statistics are plotted as time series, to show the diurnal variation of model performance. The daily statistics are plotted as bar charts to show daily performance over an episode. The macro also allows the daily results from multiple MM5 runs to be plotted together to ease the inter-comparison of performance.

STATISTICAL PERFORMANCE BENCHMARKS

Emery and co-workers (2001) have derived and proposed a set of daily performance "benchmarks" for typical meteorological model performance. These standards were based upon the evaluation of about 30 MM5 and RAMS meteorological simulations in support of air quality applications performed in the last few years, as reported by Tesche et al. (2001). The purpose of these benchmarks was not necessarily to give a passing or failing grade to any one particular meteorological model application, but rather to put its results into the proper context. For example, expectations for meteorological model performance for the U.S. west coast might not be as high as a simpler domain located over the Midwest. The key to the benchmarks is to understand how poor or good the results are relative to the universe of other model applications run for various areas of the U.S.

The statistical performance benchmarks are given in Table 3-1.

Table 3-1. Statistical benchmarks for evaluating meteorological model performance.

	Wind Speed	Wind Direction	Temperature	Humidity
RMSE	≤ 2 m/s			
Mean Bias	≤ ±0.5 m/s	≤ ±10°	≤ ±0.5 K	≤ ±1 g/kg
Gross Error		≤ 30°	≤ 2 K	≤ 2 g/kg
Index of Agreement	≤ 0.6		≤ 0.8	≤ 0.6

SURFACE STATISTICAL ANALYSES FOR WRAP

The statistical evaluation of MM5 surface fields was performed using NCAR dataset ds472, which contains hourly observations of the commonly measured variables from airports in the U.S. and Canada. Dataset variables include temperature, dew point, wind speed/direction and gusts, cloud cover fraction and cloud base for multiple cloud layers, visual range, precipitation rates and snow cover, and a descriptive weather code. The key data of interest were extracted for the various sub-domains, and processed into the appropriate formats for METSTAT.

As discussed earlier, some care must be taken in selecting an area for averaging. The problem with evaluating statistics is that the more data pairings that are summarized in a given metric, the better the statistics generally look, and so calculating a single set of statistics for a very large area (e.g., the entire 36-km domain) would not yield significant insight into performance. Therefore, a balance must be struck between taking a large enough area to create a representative sample and choosing such a large area as to smear out the signal of interest. Johnson (2003) suggests the use of the subdomains shown in Figure 3-1, and those subdomains were adopted for use in this analysis.

For each subdomain, METSTAT was used to calculate hourly and daily statistical measures, and these were compared against benchmarks for acceptable model performance set forth by Emery et al. (2001) as summarized in Table 3-1.

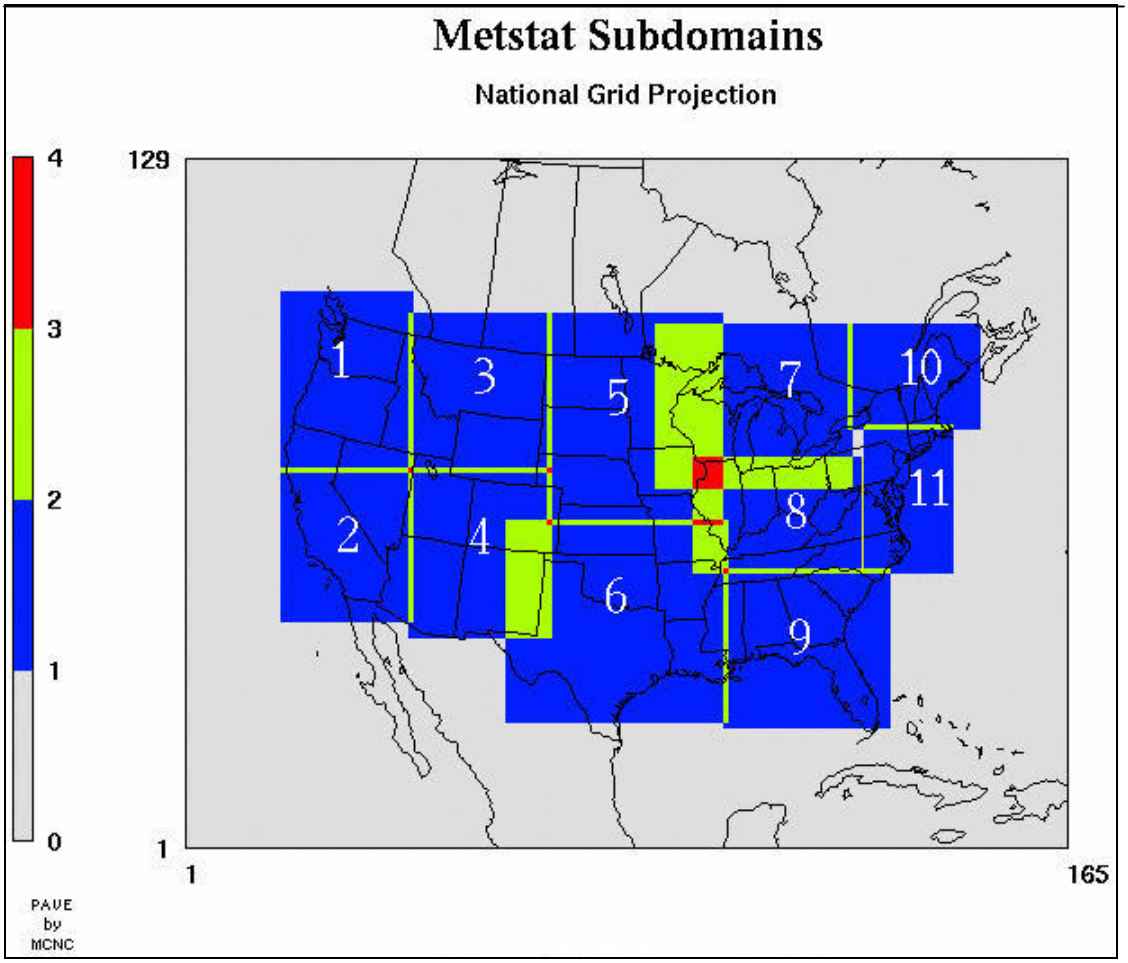

Figure 3-1. METSTAT subdomains for the WRAP 2002 MM5 performance evaluation.

PRECIPITATION EVALUATION

For each month of 2002, we developed fields of observed total rainfall over the 36 km National RPO grid and the 12 km WRAP grid. The observed precipitation amounts were generated with the CPC (Climate Prediction Center) gridded precipitation amount dataset, which is available from the National Weather Service's Climate Prediction Center at:

http://www.cpc.ncep.noaa.gov/products/precip/realtime/retro.html

The CPC daily precipitation amounts are on a latitude-longitude grid that covers the U.S. mainland at a resolution of 0.25°x 0.25°, and are ramped down to zero immediately offshore. The CPC dataset was interpolated to the 36 km and 12 km MM5 Lambert conformal grids. The advantage of the CPC precipitation field is that it has a reasonably high resolution, which is especially important when it comes to resolving the effects of orography on the precipitation over the western U.S. However, this CPC product does not include the entire 36 km or 12 km modeled domain. For example, the CPC dataset cannot be used to assess MM5's performance in simulating stratiform summer rain off the coast of California because the data effectively ends at the coastline.

4.0 SURFACE EVALUATION OF THE 2002 CENRAP, VISTAS, and WRAP RUNS

In this section, we evaluate the performance of the four 2002 36 km MM5 runs. The evaluation of the MM5 surface meteorological variables using METSTAT is presented first, followed by the evaluation of the MM5 precipitation patterns and MM5 upper-air meteorological variables.

SUBDOMAIN SURFACE FIELD METSTAT EVALUATION

The starting point for the analysis of the 2002 36 km MM5 runs was an assessment of the surface statistics for wind, temperature, and humidity using METSTAT. As described in Section 3, the continental U.S. was divided into subdomains (Figure 3-1), and each subdomain was evaluated month-by-month to isolate differences in how the model performed region by region over the course of the year. For the purpose of organizing the analysis, we will show soccerplots on which are displayed average performance statistics for each run over each subdomain for a particular month. Soccerplots are shown for wind speed RMSE versus wind direction error, temperature bias versus temperature error, and humidity bias versus humidity error. In each plot, a solid blue line indicates the benchmark. A data point that falls inside the box represents a model run that meets the performance benchmark. Perfect model performance is indicated by a data point at (0,0). The closer a data point is to the origin, the better the model's performance. It should be re-emphasized that the benchmarks are not used as an acceptance/rejection criteria of the MM5 model simulation. Rather they put the MM5 model performance into perspective and allow the identification of potential problems in the MM5 fields.

Soccerplots were generated and evaluated for each month of the year 2002. For brevity, we include only January and July in this report. MM5's performance was qualitatively different during summer and winter, with the spring and fall seasons serving as transitional periods between winter and summer. January and July were found to be representative of MM5's performance during the winter and summer seasons. The model's spring statistical performance was similar to that of fall. The main strengths and weaknesses of the 2002 MM5 runs are captured in the January and July plots. For both months, we show a soccerplot for the subdomains in the western U.S. (the WRAP region) and a second soccerplot for the central and eastern U.S. This is done to reduce the amount of data shown on one plot to a manageable level, and also because MM5's performance was qualitatively different in the western U.S. and the rest of the country.

Table 4-1 lists the METSTAT domains shown in Figure 3-1 and corresponding abbreviations used in the discussion below.

Table 4-1. METSTAT subdomain abbreviations (see Figure 3-1).

METSTAT Subdomain	Name Abbreviation
1	PacificNW
2	SW
3	North
4	DesertSW
5	CenrapN
6	CenrapS
7	Great Lakes
8	Ohio Valley
9	SE
10	NE
11	MidAtlantic

January MM5 Surface Performance

Comparison of the wind soccerplots for January (Figures 4-1 and 4-2) shows a marked division by geographic region, with the western subdomains PacificNW, SW, DesertSW, North, and NE falling outside the benchmarks for the CENRAP, VISTAS, and WRAP_0 runs. In the eastern and central U.S., wind performance is within the benchmark for all four runs for all subdomains except NE. In the final WRAP run, the western subdomains are within or close to (SW) the benchmark. Performance in all central and eastern subdomains improved, and the NE subdomain is brought closer to the benchmark. The final WRAP run is the best performing run overall for surface wind in January.

Figures 4-3 and 4-4 show January time series for the initial and final WRAP runs for wind speed and wind speed bias and wind direction and wind direction bias for the desertSW subdomain. The WRAP_0 run reproduced the overall shape of the observed wind speed time series, but had a significant low wind speed bias. In the final WRAP run, this bias is reduced to the point where the desertSW subdomain falls within the performance benchmark for wind speed. There is also an improvement in simulation of wind direction. The WRAP_0 run has a positive wind direction bias throughout much of January, and this bias is reduced in the final WRAP run (Figure 4-4). The final WRAP run also did a better job of reproducing the large wind direction excursions on January 12 and January 30. In the final WRAP run, there is an overall reduction in wind speed RMSE relative to the WRAP_0 run (time series not shown) over the entire month. The improvement in the simulation of surface winds in the final WRAP run likely occurs because of nudging of surface winds to observations. Surface obs nudging of winds is not done in the WRAP_0 runs, nor in the CENRAP or VISTAS runs.

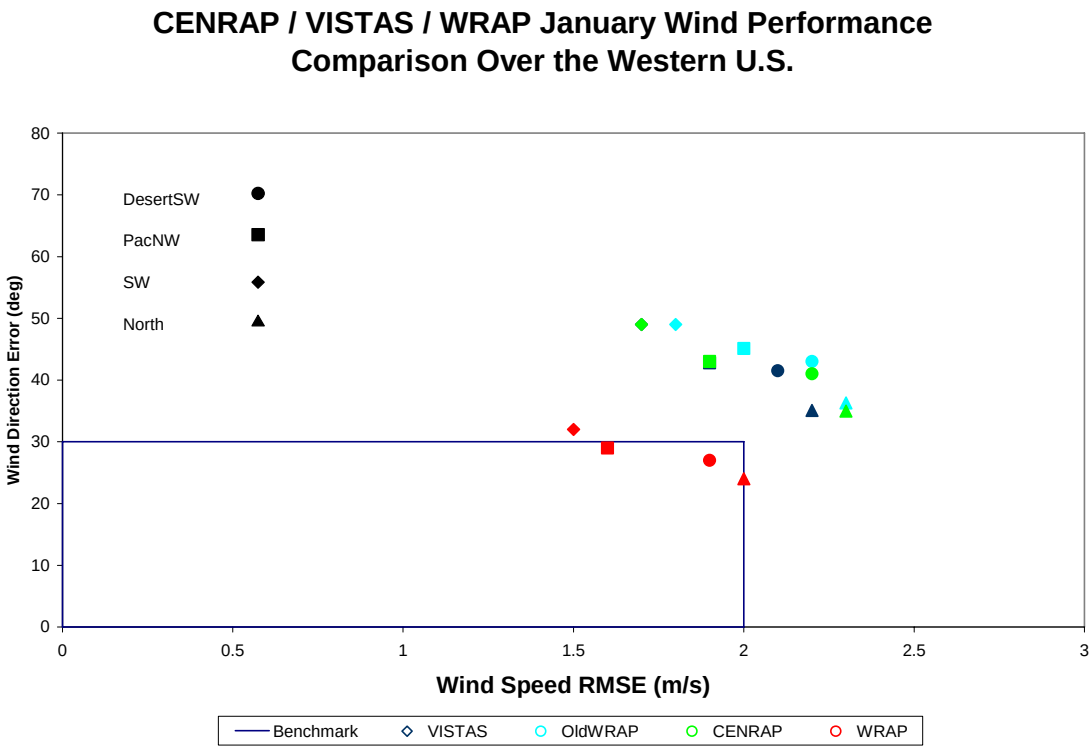

Figure 4-1. 36 km surface wind comparison soccerplot for January over western U.S. for VISTAS, CENRAP, WRAP_0 (old WRAP) and final WRAP (WRAP) 2002 36 km simulation.

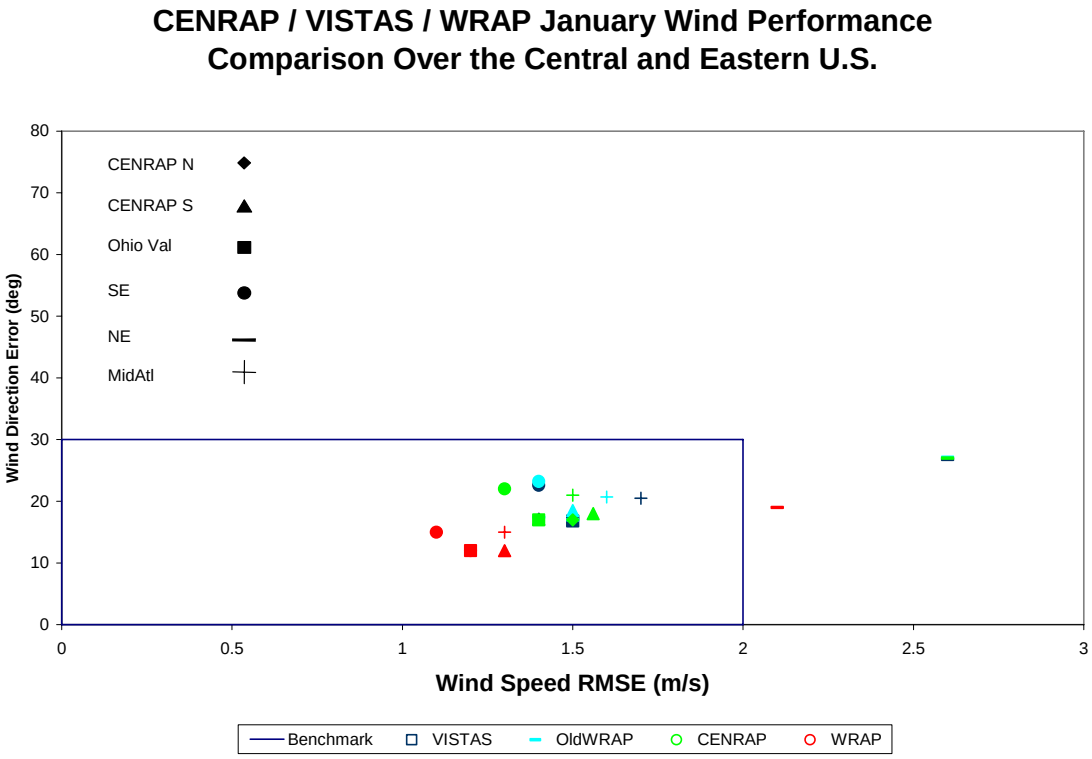

Figure 4-2. 36 km surface wind comparison soccerplot for January over central and eastern U.S. for VISTAS, CENRAP, WRAP_0 (old WRAP) and final WRAP (WRAP) 2002 36 km simulation.

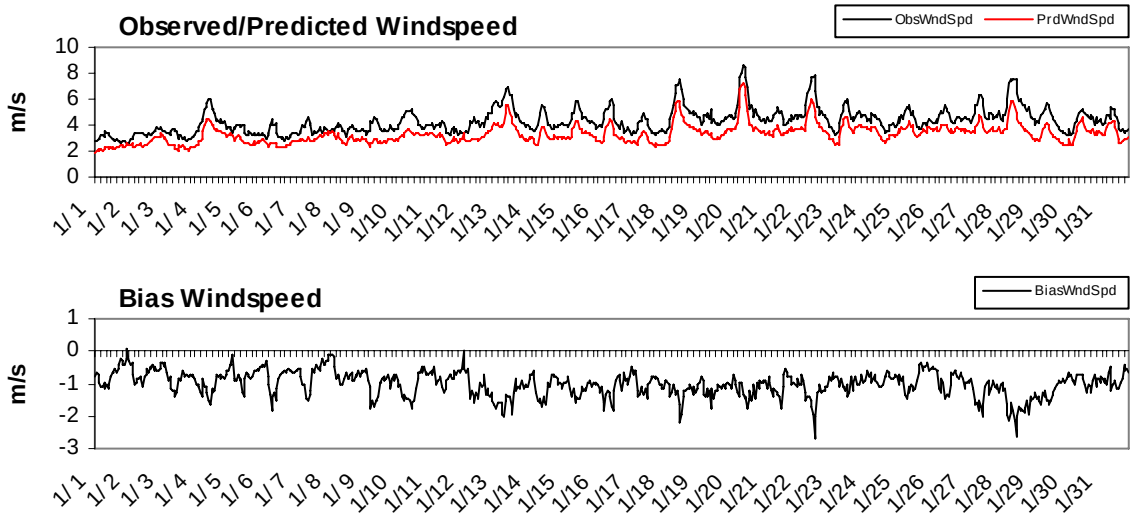

Figure 4-3. WRAP_0 run surface wind speed and bias for January over desertSW subdomain.

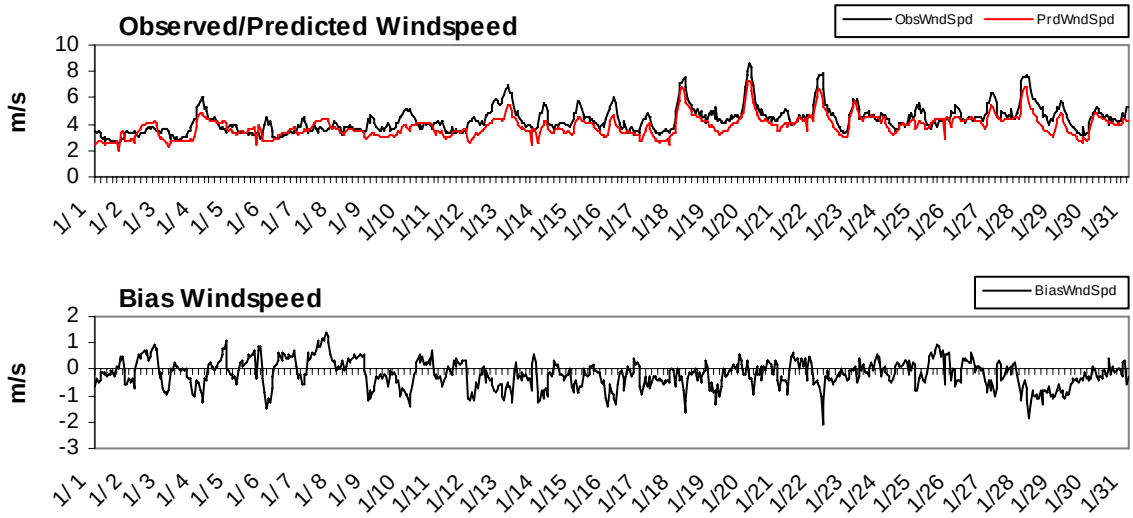

Figure 4-4. Final WRAP run surface wind speed and bias for January over desertSW subdomain.

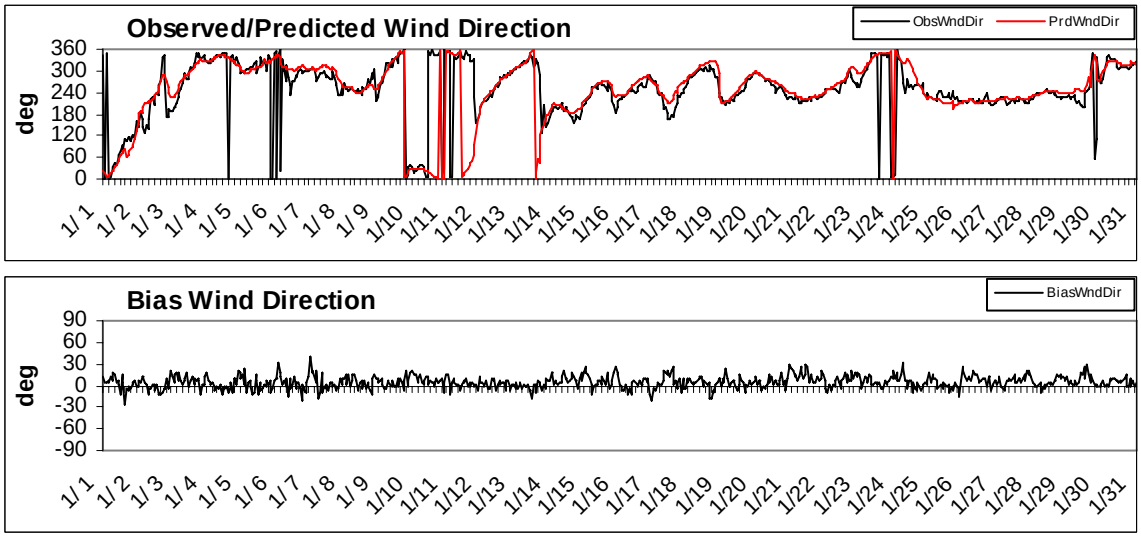

Figure 4-5. WRAP_0 run surface wind direction and bias for January over desertSW subdomain.

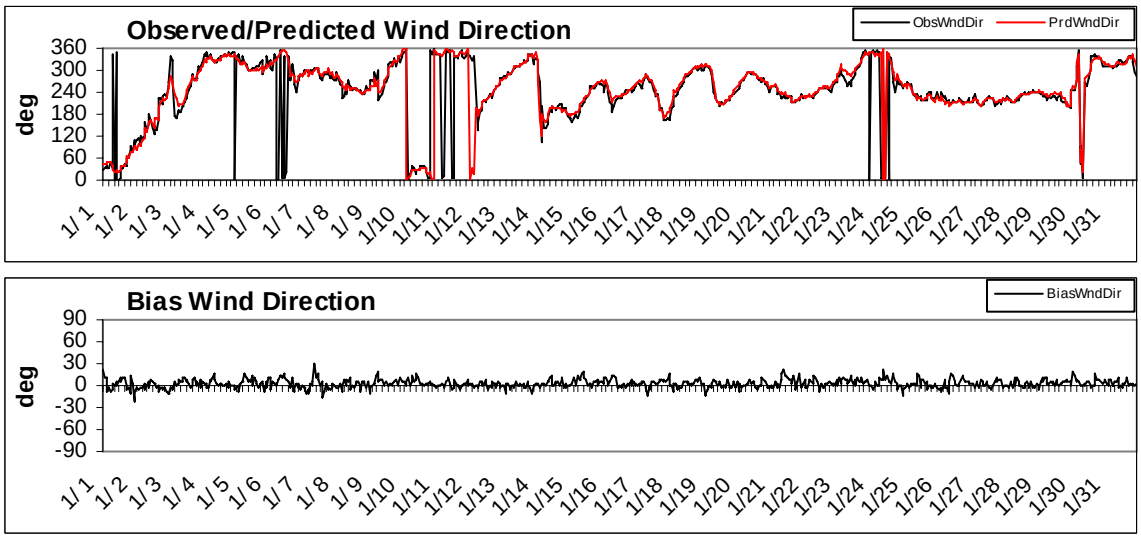

Figure 4-6. Final WRAP run surface wind direction and bias for January over desertSW subdomain.

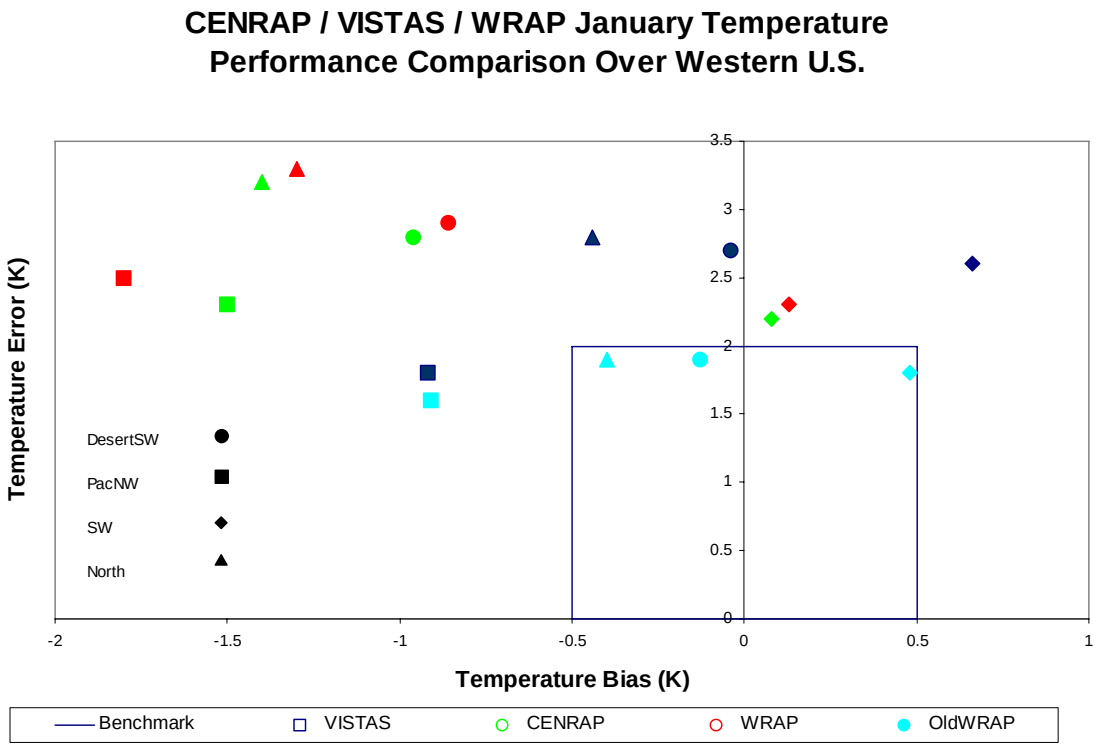

Figure 4-7. 36 km temperature comparison soccerplot for January over western U.S. for VISTAS, CENRAP, WRAP_0 (old WRAP) and final WRAP (WRAP) 2002 36 km simulation.

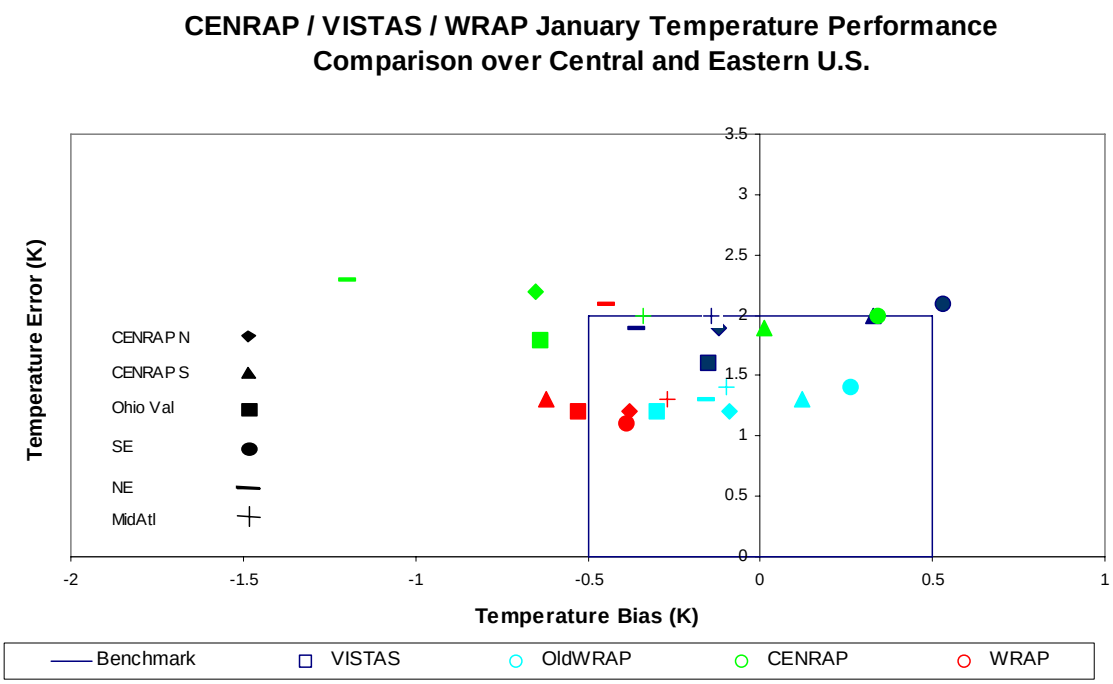

Figure 4-8. 36 km temperature comparison for January over central and eastern U.S. for VISTAS, CENRAP, WRAP_0 (old WRAP) and final WRAP (WRAP) 2002 36 km simulation.

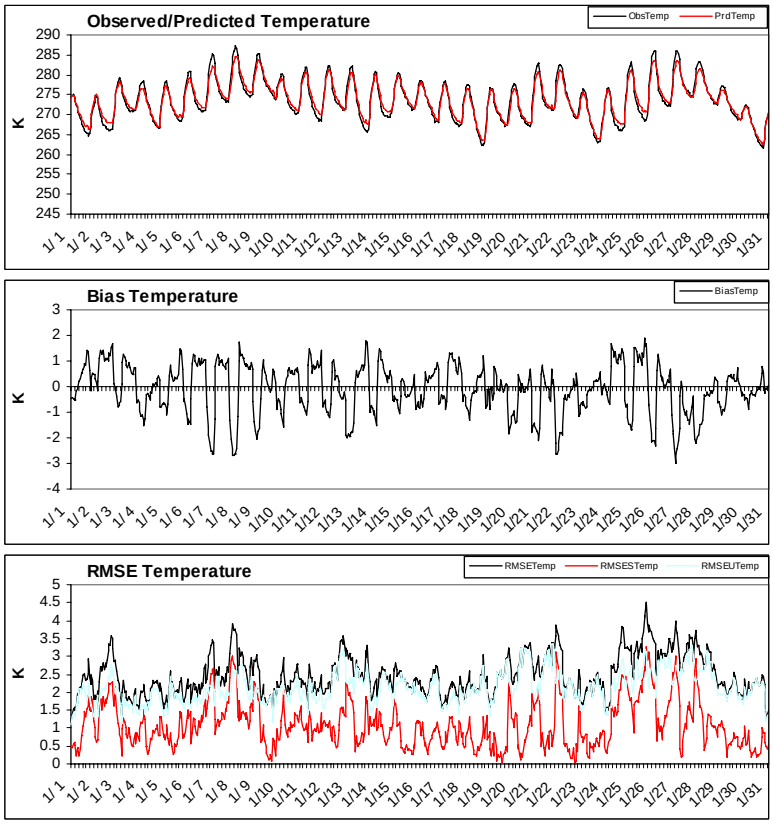

Figure 4-9. WRAP_0 run January temperature time series for desertSW subdomain.

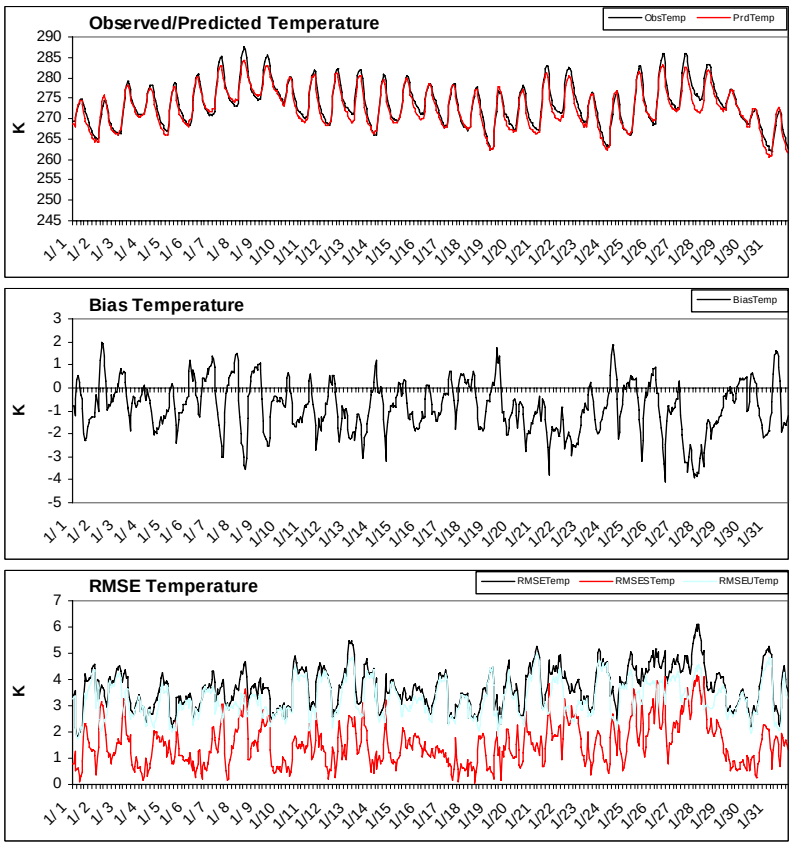

Figure 4-10. Final WRAP run January temperature time series for desertSW subdomain.

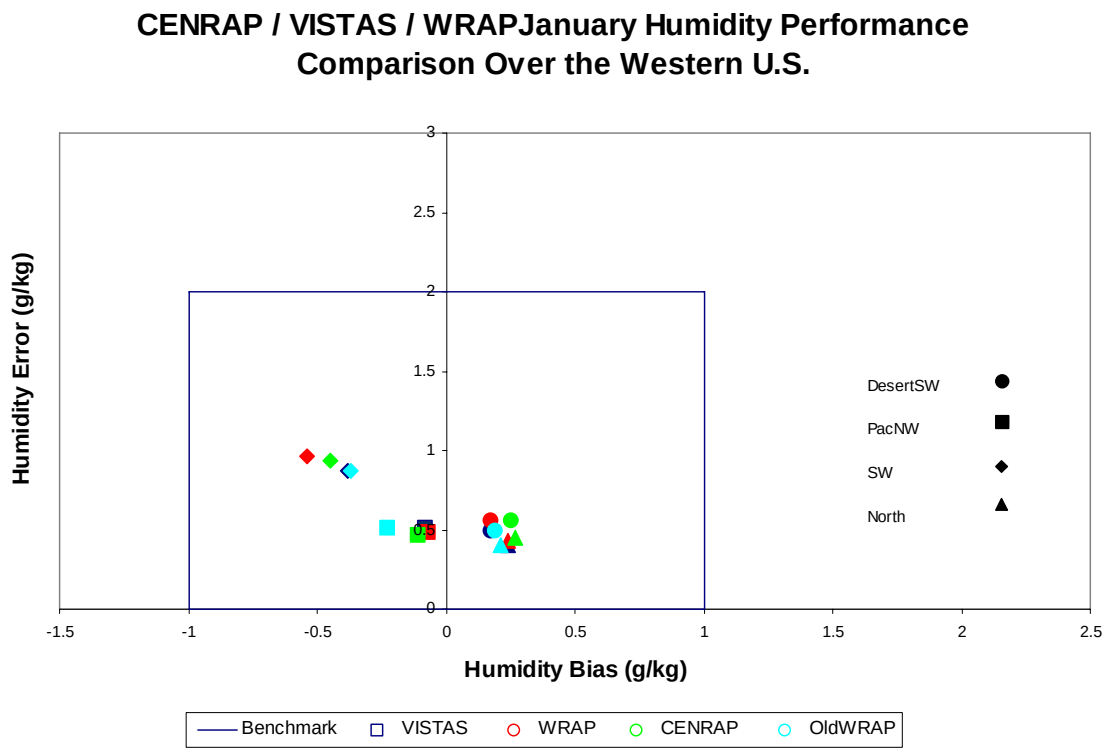

Figure 4-11. 36 km humidity comparison soccerplot for January over western U.S. for VISTAS, CENRAP, WRAP_0 (old WRAP) and final WRAP (WRAP) 2002 36 km simulation.

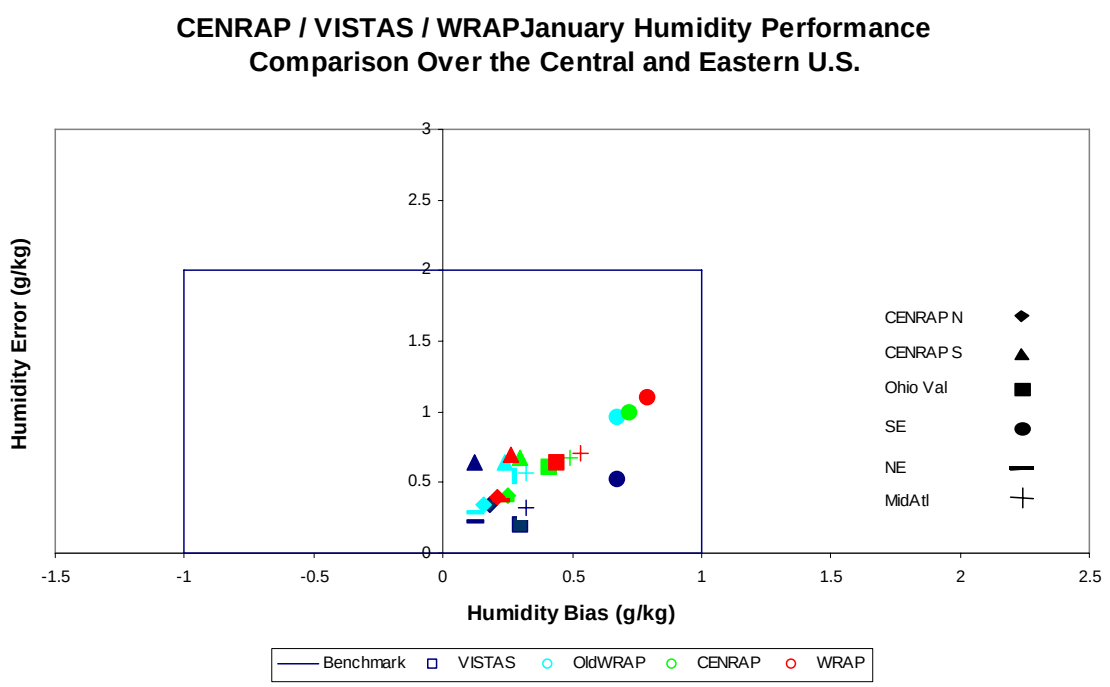

Figure 4-12. 36 km humidity comparison soccerplot for January over central and eastern U.S. for VISTAS, CENRAP, WRAP_0 (old WRAP) and final WRAP (WRAP) 2002 36 km simulation.

For temperature in January (Figures 4-7 and 4-8), the WRAP_0 run outperformed the other runs, with three of four western subdomains and all eastern and central subdomains in the goal. This was, of course, because surface analysis nudging of temperature was performed in the WRAP_0 run, but not the other three runs. With the analysis nudging turned off, the performance of the final WRAP run becomes comparable to the CENRAP run in the western subdomains. Of the three runs with no surface analysis nudging of temperature (VISTAS, CENRAP, final WRAP), VISTAS is the best-performing run. Figure 4-9 and Figure 4-10 show the change in surface temperature statistic time series from WRAP_0 to the final WRAP run for the desertSW subdomain. Over the month of January, the bias and the RMSE both increase in the final WRAP run (note that the RMSE is on a different scale in the old and final WRAP runs). The final WRAP run has a larger cold bias than the WRAP_0 run for the desertSW subdomain, and this increased cold bias was common to all subdomains across the U.S. in going from the WRAP_0 run to the final WRAP run

In the eastern and central subdomains, the temperature error does not change significantly between the WRAP_0 and final WRAP runs (except for NE), but the cold bias increases (Figure 4-8). The cause of this is unclear. Overall, most subdomains show a cold bias. In the central and eastern subdomains, all four runs tended to be within or near the benchmark, but in the west, all runs were outside the benchmark except for the analysis nudged WRAP_0 run. Olerud and Sims (2003) have suggested that spurious temperature differences are introduced by mismatches between observing station elevation and modeled terrain heights that is caused in part by coarse model grid resolution used.

For January humidity (Figures 4-11 and 4-12), there is a general wet bias except in the SW and PacNW subdomains. Differences among the runs are small, except that VISTAS does better in the Ohio Valley, SE and MidAtlantic subdomains. All four runs lie within the benchmark for all subdomains.

Overall for January, all four model runs tended to be too wet, too cold, and had a low wind speed bias (not shown). The surface analysis nudged WRAP_0 run performed best for temperature, with the VISTAS run the next best performer. All four runs performed better in the central and eastern subdomains for wind and temperature, and this is likely due to the effects of modeling the topography of the west at 36 km resolution. There was not much difference among the runs in humidity performance in the west, but VISTAS was clearly the best performing run in the east.

July MM5 Surface Model Performance

In July, the wind direction error and the wind speed RMSE (Figure 4-13 and Figure 4-14) show a general increase relative to January (Figures 4-1 and Figure 4-2) in the western and central and eastern subdomains. In the western subdomains, there is not much difference in performance among the VISTAS, CENRAP, and WRAP_0 runs, although VISTAS does slightly better than the others. The final WRAP run shows a slight decrease (improvement) in wind speed RMSE and a larger decrease in wind direction error relative to the other runs. Again, this is due to nudging to surface wind observations, which is done in the final WRAP run, but not in the VISTAS, CENRAP, or WRAP_0 runs. As in January, MM5 does better in simulating winds in the central and eastern U.S. than it does in the west. In the central and eastern subdomains, there is not much difference among the WRAP_0, CENRAP, and VISTAS runs in terms of July wind performance. The final WRAP run, which is nudged to surface wind observations, displays improved wind direction error relative to the other runs, as well as a smaller wind speed RMSE.

For temperature (Figures 4-15 and 4-16), July performance is better in the east than the west, as seen for wind. The temperature RMSE is generally smaller in the east, and the western subdomains have a much stronger cold bias. As mentioned earlier, part of this cold bias in the west may be due to terrain resolution effects. In the eastern and central U.S. (Figure 4-16), temperature performance among the WRAP_0, CENRAP, and VISTAS runs is roughly equivalent, while the final WRAP run has an outlier in the cenrapS subdomain. In the west, CENRAP is the best performing run, as it is the only one of the four runs with no outlier. The final WRAP run has three of four subdomains within the benchmark for bias, but its SW subdomain has a strong warm bias, and is the only western subdomain in any of the four runs to have a warm bias. All runs except the final WRAP run showed large cold biases in the desertSW subdomain. This is likely due to excessive simulated convective rainfall, which affects the partitioning of surface energy, diverting an excess of energy to latent rather than sensible heating. This lowers the daily surface temperature maximum (Kemball-Cook et al., 2004a). For example, Figures 4-17 and 4-18 show the time series of temperature and temperature bias for the SW subdomain. The WRAP_0 run has a cold bias caused by a consistent underestimation of the daily maximum surface temperature. Precipitation was overestimated in the SW subdomain in the WRAP_0 run. In the final WRAP run, the rainfall is much closer to observations than in WRAP_0, and the cold bias has shifted to a warm bias. The North subdomain (Figures 4-19 and 4-20) exhibits similar behavior.

For humidity, (Figure 4-21), the error is higher in July than in January for all subdomains. Over the central and eastern U.S., performance among the WRAP_0 CENRAP, and VISTAS runs is roughly comparable. The final WRAP run has a cold bias for most subdomains, and outliers cenrapN and cenrapS. In the western U.S. the WRAP_0 run has outliers desertSW and North, which have a strong wet bias related to excess precipitation over these subdomains. This strong wet bias over the WRAP region was one of the main motivations for redoing the WRAP run, and the final WRAP run does show improved humidity performance in July, with all western subdomains now lying within the benchmark. The cost of this improvement in WRAP humidity performance in the west is degraded humidity performance in the east.

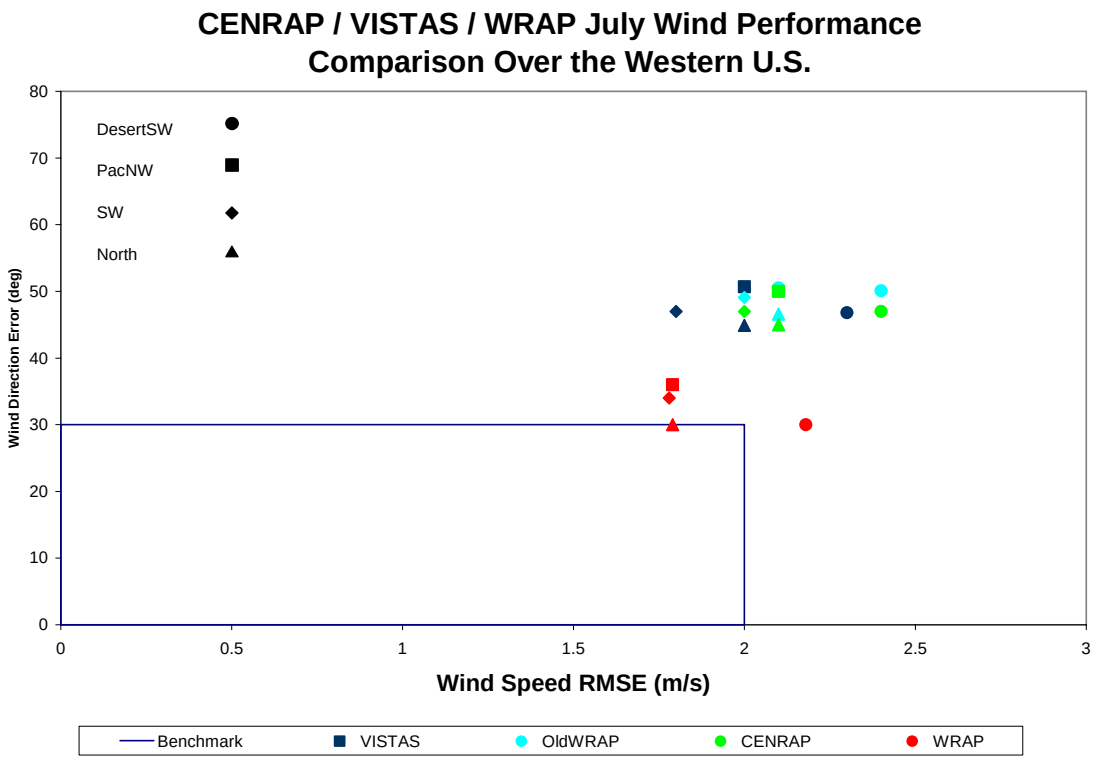

Figure 4-13. 36 km wind comparison soccerplot for July over western U.S. for VISTAS, CENRAP, WRAP_0 (old WRAP) and final WRAP (WRAP) 2002 36 km simulation.

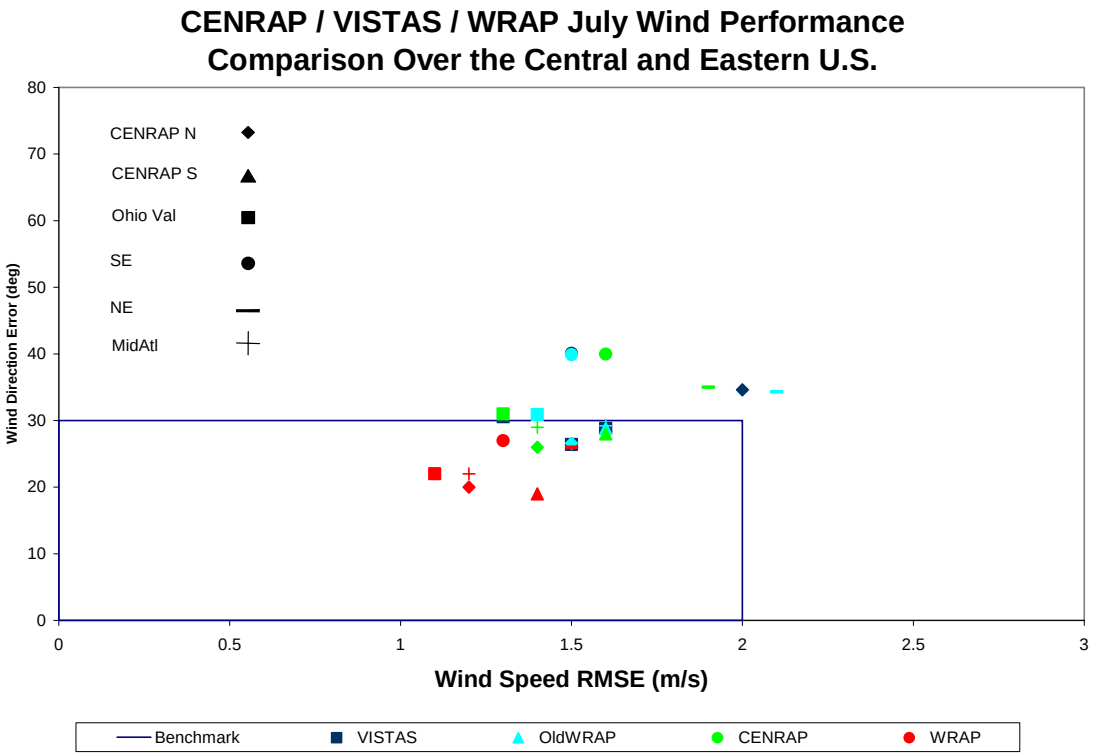

Figure 4-14. 36 km wind comparison soccerplot for July over central and eastern U.S. for VISTAS, CENRAP, WRAP_0 (old WRAP) and final WRAP (WRAP) 2002 36 km simulation.

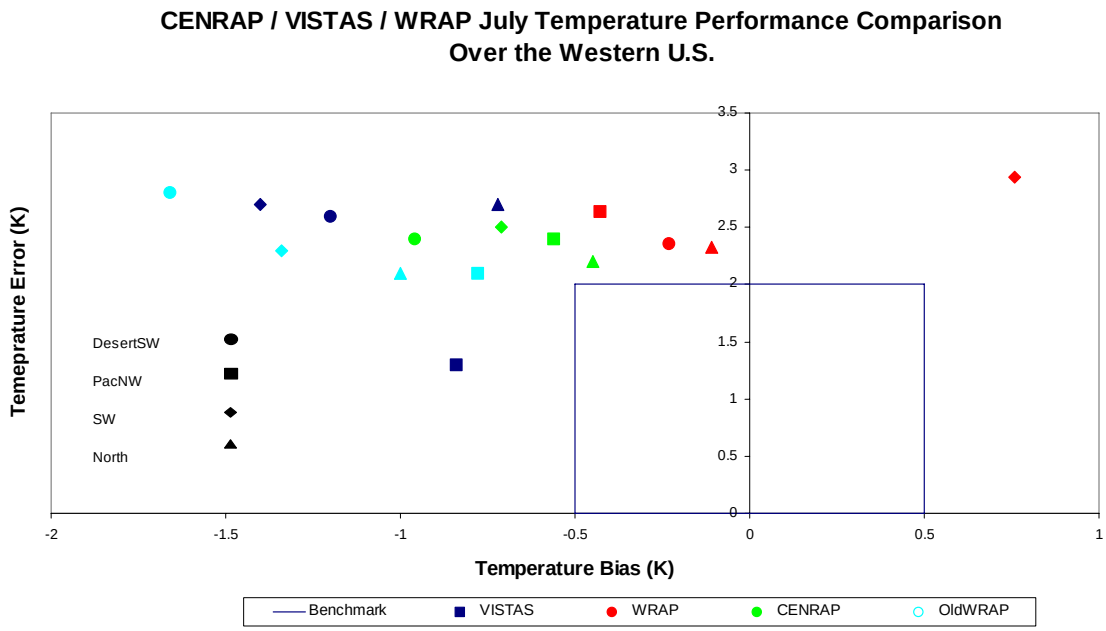

Figure 4-15. 36 km temperature comparison soccerplot for July over western U.S. for VISTAS, CENRAP, WRAP_0 (old WRAP) and final WRAP (WRAP) 2002 36 km simulation.

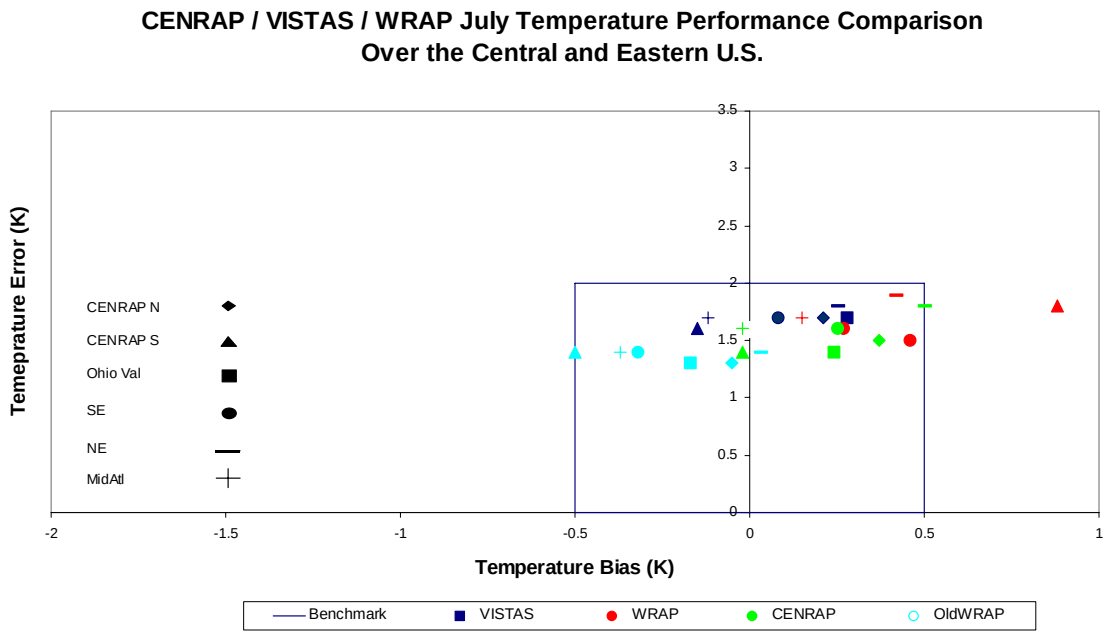

Figure 4-16. 36 km temperature comparison soccerplot for July over central and eastern U.S. for VISTAS, CENRAP, WRAP_0 (old WRAP) and final WRAP (WRAP) 2002 36 km simulation.

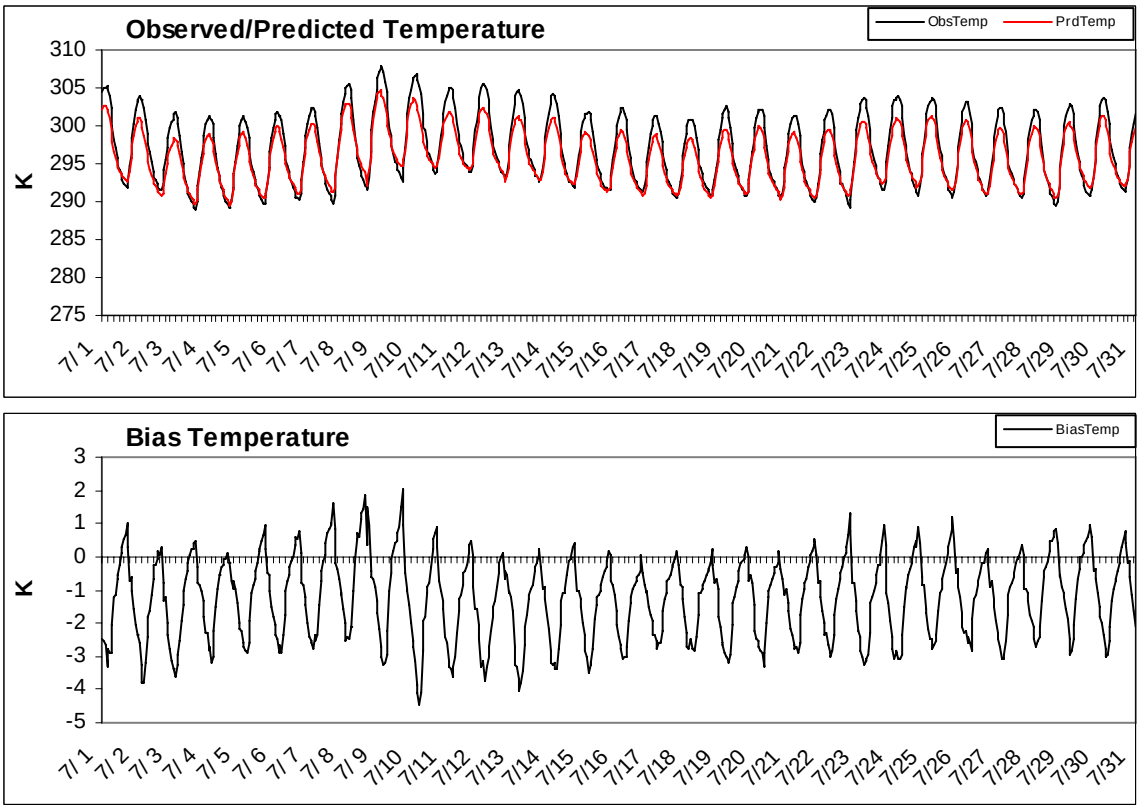

Figure 4-17. WRAP_0 run July temperature time series for SW subdomain.

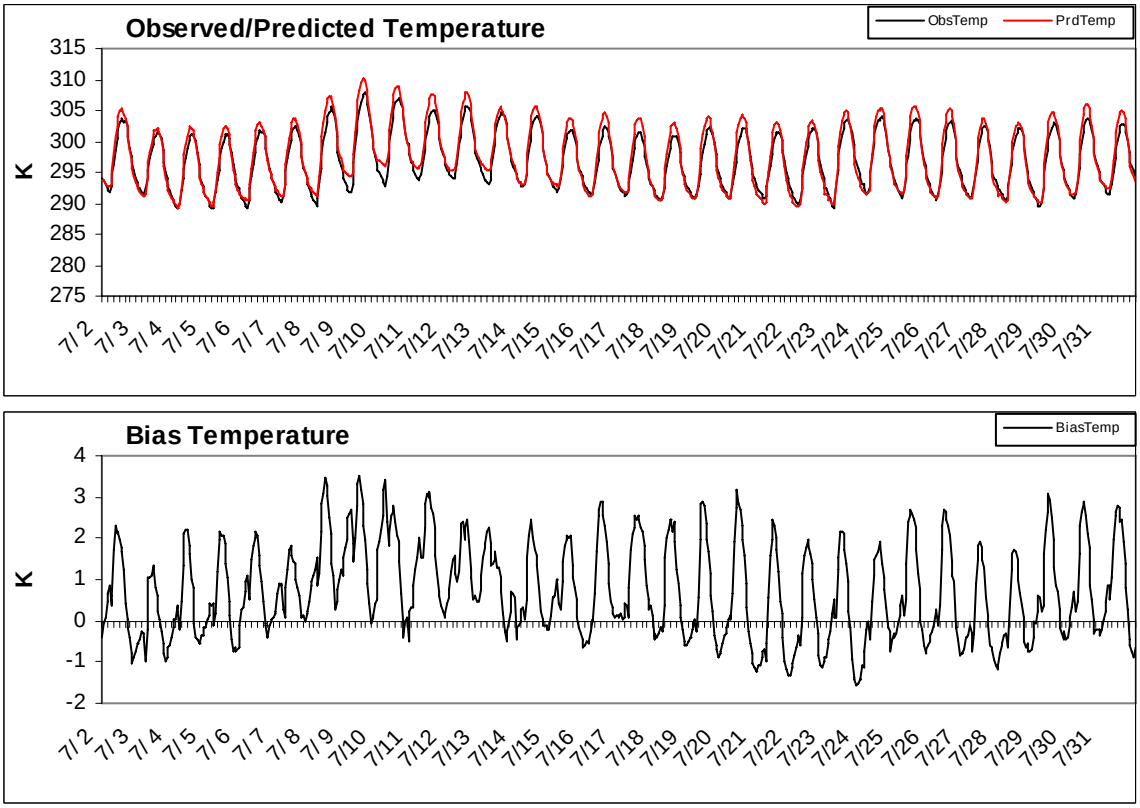

Figure 4-18. Final WRAP run July temperature time series for SW subdomain.

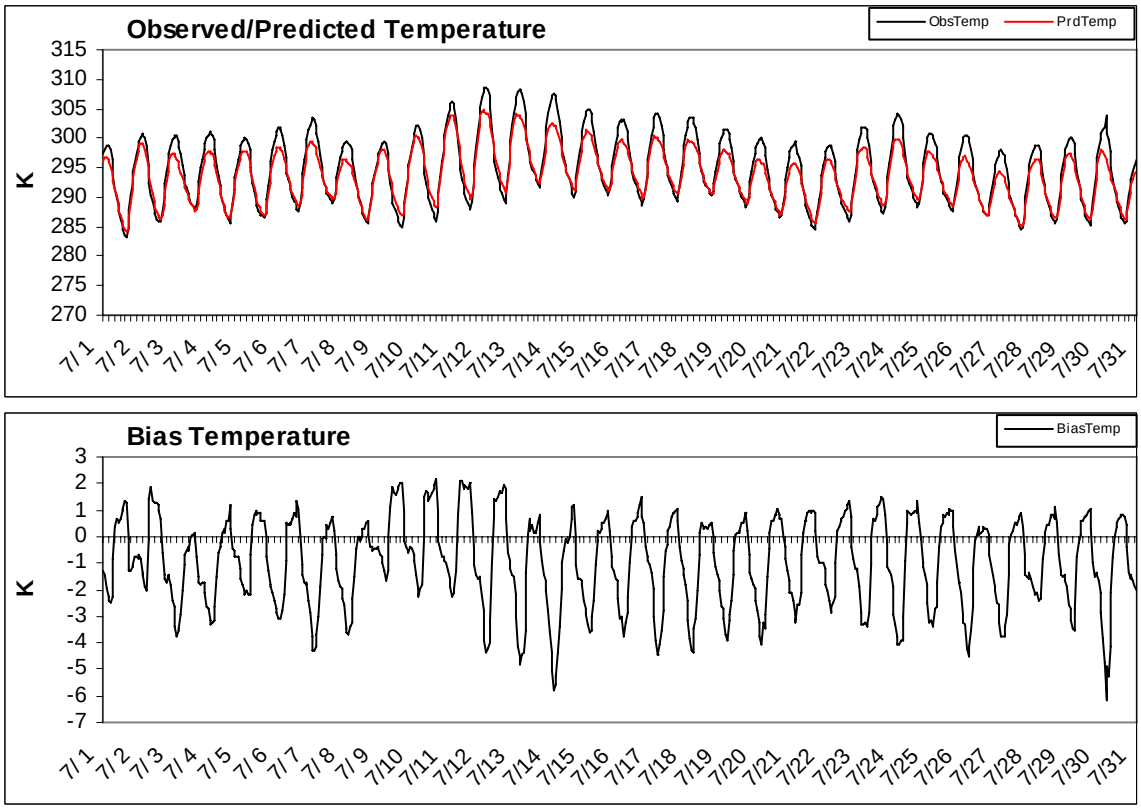

Figure 4-19. WRAP_0 run July temperature time series for North subdomain.

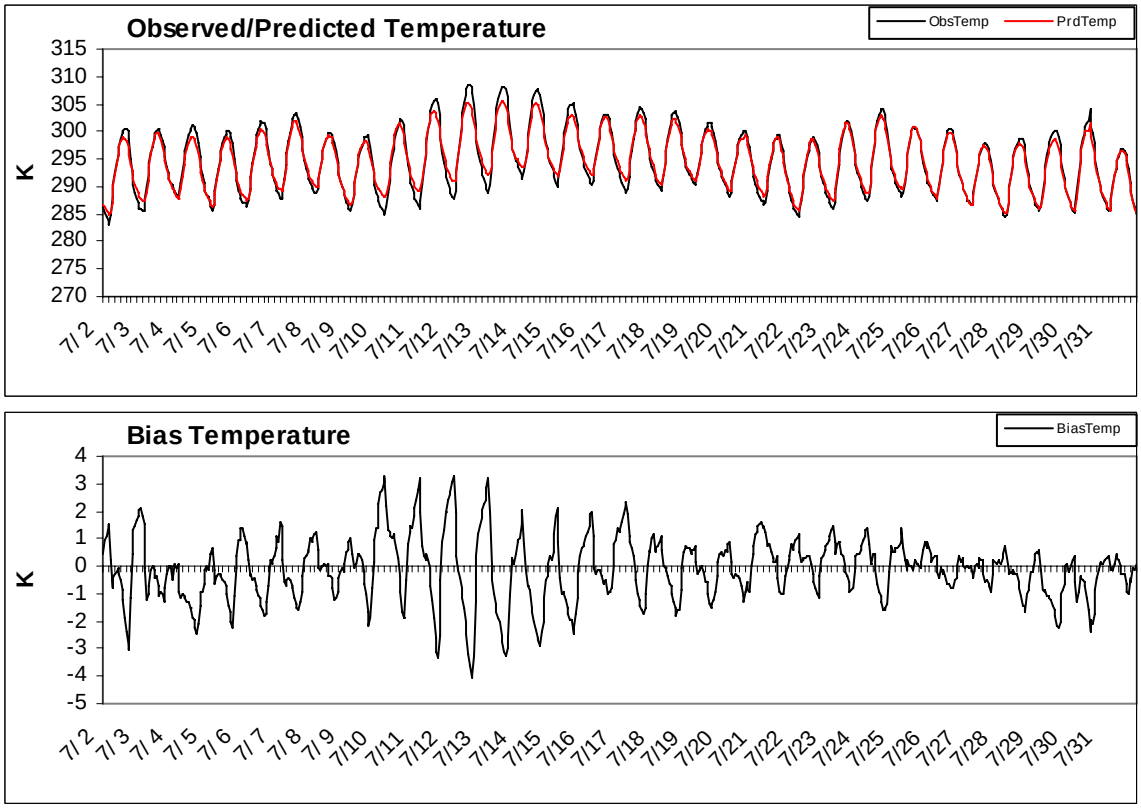

Figure 4-20. Final WRAP run July temperature time series for North subdomain.

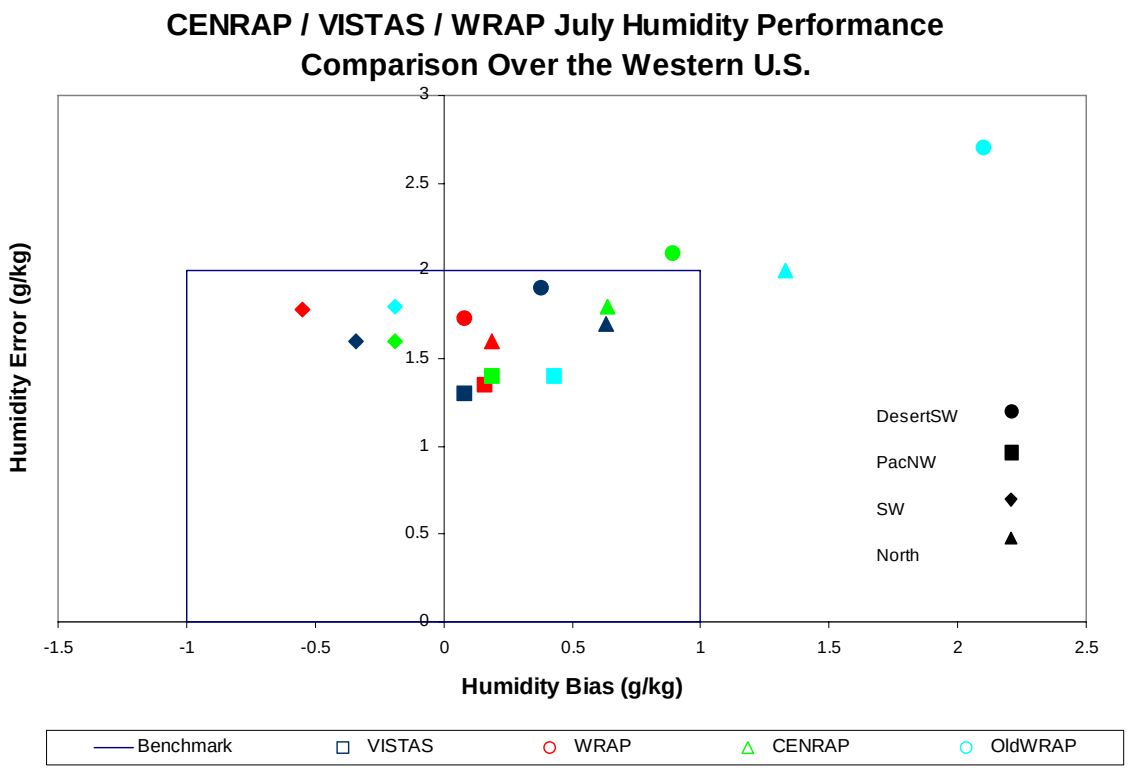

Figure 4-21. 36 km humidity comparison soccerplot for July over western U.S.

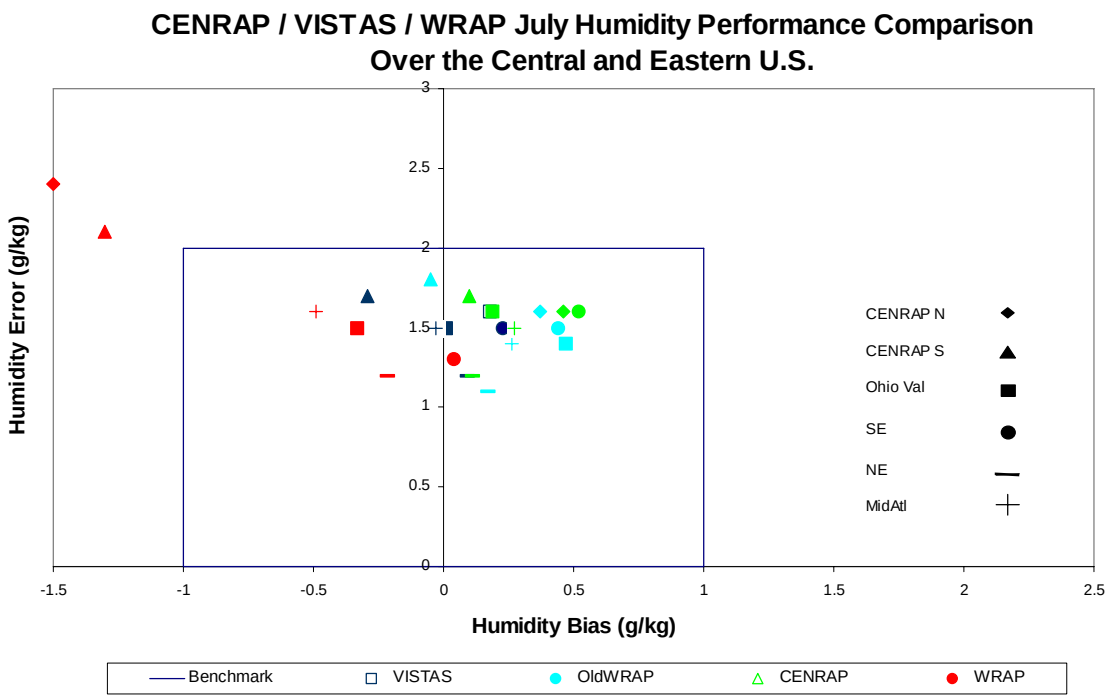

Figure 4-22. 36 km humidity comparison soccerplot for July over eastern and central U.S.

Summary of the Annual MM5 Surface Performance

- For temperature and humidity, the best performing run overall was VISTAS because it had no outliers and has no serious performance problems at any particular time or subdomain.

- For winds, the best performing run was the final WRAP run. There was not much difference in wind performance among the other three runs.

- For wind, there is not much variation in performance in the east or the west over the course of the annual cycle, although the wind direction error is smaller in winter than in summer. The wind direction RMSE does not vary significantly over the course of the year for any of the four runs.

- For temperature in the west, all runs except the nudged WRAP_0 run lie outside the benchmark (except for the SW and PacificNW subdomains, which lie barely within the benchmark in December). Performance is much better in the east than in the west.

- The final WRAP run had a less accurate surface temperature simulation than WRAP_0, which used analysis nudging of surface temperature

- For temperature in the east, all or nearly all of the subdomains lie within the benchmark during most months.

- Humidity performance was generally within the benchmarks for all regions for all runs except the WRAP_0 run.

- For humidity, in the west, performance is best in winter and deteriorates as summer approaches. The culprit is the humidity error, not the bias, except for the WRAP_0 run outliers in July in north and desertSW.

- Except for SW, all western subdomains have a wet bias in July. This is likely related to excess simulated precipitation.

- For eastern subdomains, humidity performance is comparable to western subdomains, as for wind.

- The final WRAP run showed significant improvement in July humidity performance in the west relative to the WRAP_0 run, but does worse in the east. We can attribute both of these changes in performance to reduction in rainfall and therefore surface moisture that we get in going from the Kain-Fritsch cumulus scheme to the Betts-Miller scheme.

EVALUATION OF PRECIPITATION IN THE 36 KM 2002 CENRAP, VISTAS, AND WRAP RUNS

In this section, we evaluate the precipitation fields in the VISTAS, CENRAP, WRAP_0 and final WRAP 2002 36 km MM5 runs. We examine the annual cycle in the monthly rainfall totals in the MM5 runs and compare them to observed monthly rainfall totals.

In January (Figure 4-23), the rainfall amount is predicted with good skill in all four runs. All four runs overestimate rainfall in the cenrapN and North subdomains. The final WRAP run produces more rainfall than the other three runs in the mountains along the west coast; this brings the final WRAP run into closer agreement with observations than the other three runs in this region. This rainfall increase is due to the stratiform rainfall component (Figure 4-24), and is associated with mid-latitude cyclones coming in off the Pacific Ocean. The convective component of the rainfall decreases slightly over the western coastal regions in the final WRAP run (Figures 4-24). The surface humidity for the PacificNW subdomain for January (Figure 4-11) shows an improvement in the surface humidity in going from the WRAP_0 to the final WRAP run, which is reasonable, given the more realistic rainfall simulation.

In March, rainfall is predicted with good skill in all four runs (Figure 4-25). All four runs underpredict the maximum over Arkansas and translate it to the east. All four runs overpredict rainfall over the desertSW subdomain. Comparing the new and old WRAP runs (Figure 4-25), we see that there is a signal of increased rain over the storm track areas, but it is smaller than the increase noted above for January. The area of increased intensity over the southeastern US (Kentucky) in the final WRAP run is an improvement over WRAP_0.

In July (Figure 4-26), rainfall is overpredicted in all four runs over most of the U.S. This is a result of excessive convective rainfall (e.g., Figure 4-27). Results for CENRAP and VISTAS runs are similar (but are not shown). The problem is most severe in the WRAP_0 run, and is somewhat ameliorated in the final WRAP run. The overprediction of rainfall (especially in the North and desertSW subdomains) is associated with the wet humidity bias seen in the surface soccer plots (Figure 4-21). The overprediction of rainfall in the southwestern U.S. is smallest in the final WRAP run, which has a different type of cumulus scheme than the other three runs. Overall, the level of skill could be improved in all four runs.

Comparing July precipitation in the new and old WRAP runs, we see a large reduction in the number of grid cells experiencing convective rainfall in the final WRAP run. This is the main reason for the improvement in the surface humidity in the desertSW and North subdomains. However, in regions where convection did occur, it tended to be overly strong in the final WRAP run. Note, for instance, the increase in wet bias in the SE subdomain in Figure 4-22. Inspection of the July precipitation plot for the final WRAP run shows that the precipitation field is very granular in regions with light precipitation. It is not uncommon to have a grid cell with no rain bordering a grid cell where heavy rain has fallen. This behavior is not observed in any of the other three runs, which use Kain-Fritsch schemes. This suggests that it is more difficult to trigger convection in the Betts-Miller scheme than in the Kain-Fritsch scheme, which shows a larger number of convecting grid cells which rain weakly.

Our findings in this study are similar to those of Gochis et al. (2002), who studied the sensitivity of the simulation of North American Monsoon precipitation in MM5 to the choice of convective scheme. Gochis et al. (2002) noted:

> "The trigger function in the Betts-Miller-Janjic scheme also appears to inhibit convective activity in the northern North American Monsoon regions, which gives reduced precipitation and a relatively cooler and drier mid level atmosphere compared with the Kain-Fritsch formulation…Where Betts-Miller-Janjic is activated… there appears to be an overestimate of precipitation. This suggests that the profile adjustment procedure used in the Betts-Miller-Janjic scheme is either yielding too much column water during the relaxation of the large-scale profile toward the reference profile, or that too much of the residual moisture is being converted into precipitation."

In their study, the MM5 runs using the Kain-Fritsch scheme tended to produce rainfall that was more widespread than did MM5 running with the Betts-Miller scheme, but when the Betts-Miller scheme did switch on, it tended to rain more heavily than Kain-Fritsch. Their hypothesis is that the Kain-Fritsch scheme is more easily triggered than the Betts-Miller scheme, and they support this hypothesis by showing that the Kain-Fritsch scheme consistently turned on earlier in the day than the Betts-Miller scheme.

In October (Figure 4-28), all four runs underpredict precipitation in the cenrapN, cenrapS and NE subdomains, and overpredict in the Ohio Valley subdomain. All three runs have excessive precipitation over Florida, and underpredict and/or misplace the maximum over Louisiana. Comparison of the new and old WRAP runs shows a reduction in the area undergoing convective rainfall in the Gulf of Mexico and Atlantic Ocean in regions where the water is still warm. Where convection occurs, however, it tends to be more intense in the final WRAP run than in the WRAP_0 run.

The December precipitation fields (Figure 4-29) are similar to those of January (Figure 4-23). In December, all four models predict the precipitation field with good skill, although the maximum over Arkansas/Louisiana is underestimated and the intensity of the precipitation over the Oregon coast is too low.

In summary, differences among the four runs in terms of precipitation performance were relatively small. All four runs were able to reproduce many of the major stratiform (i.e. resolved at grid-scale) precipitation features, and all four had difficulty with convective precipitation. This is to be expected, given that convection is a subgrid-scale phenomenon whose parameterization is poorly understood (e.g. Randall et al. 2003). The biggest difference among the four simulations came in July; in the western U.S., the overestimation of precipitation was smallest in the final WRAP run, but the final WRAP run also overpredicts convective rainfall in the southeastern U.S. These changes are clearly related to the use of a different type of cumulus scheme in the final WRAP run than in the other three runs. For the purposes of WRAP visibility modeling, the second WRAP run is clearly an improvement over the WRAP_0 and other runs, and has more accurate precipitation fields in the western U.S. in both summer and winter.

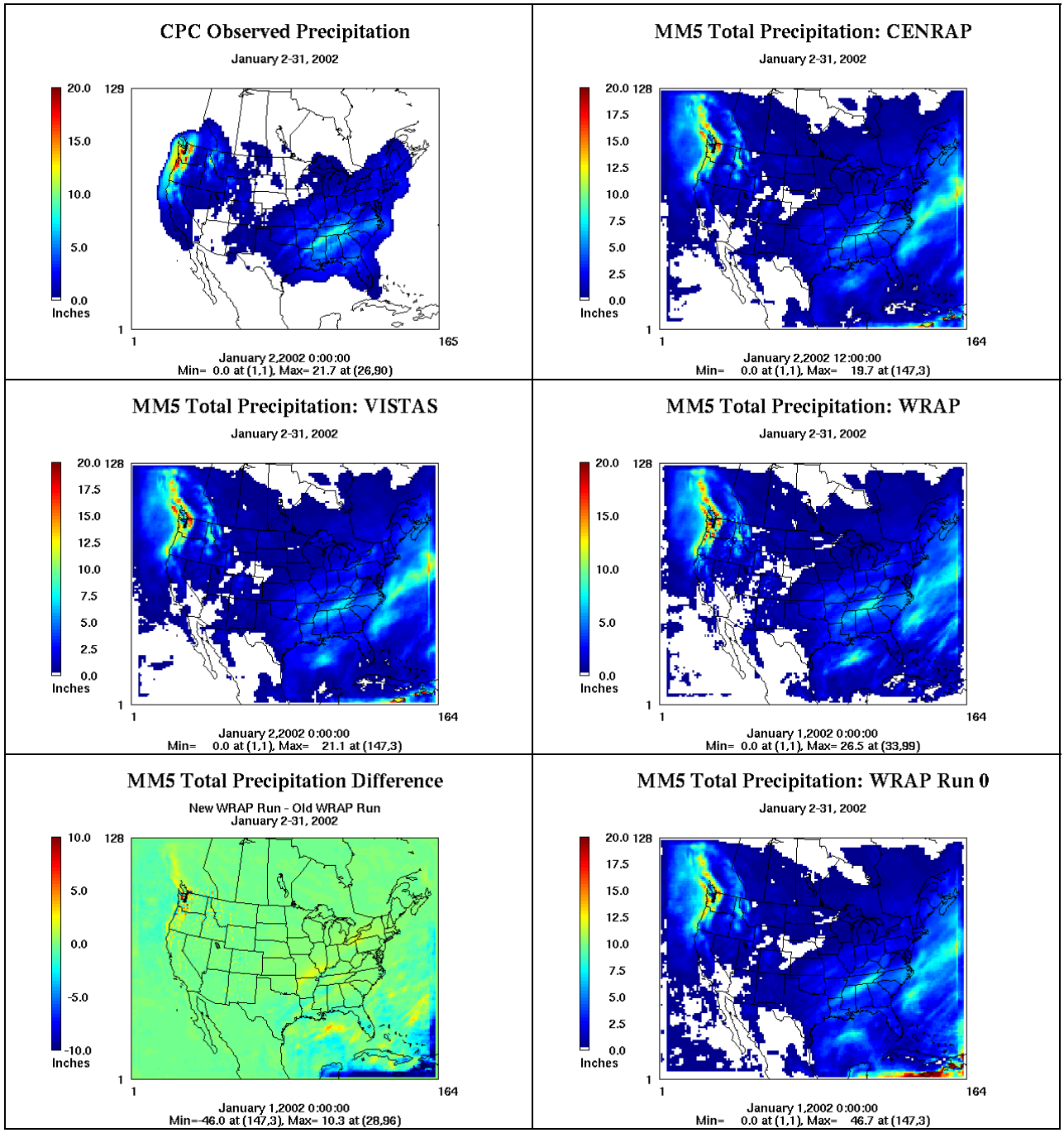

Figure 4-23. January observed (top left) and modeled precipitation for CENRAP (top right), VISTAS (middle left), final WRAP (middle right) and WRAP_0 (bottom right) 2002 36 km simulations and differences between WRAP-0 and final WRAP (bottom left).

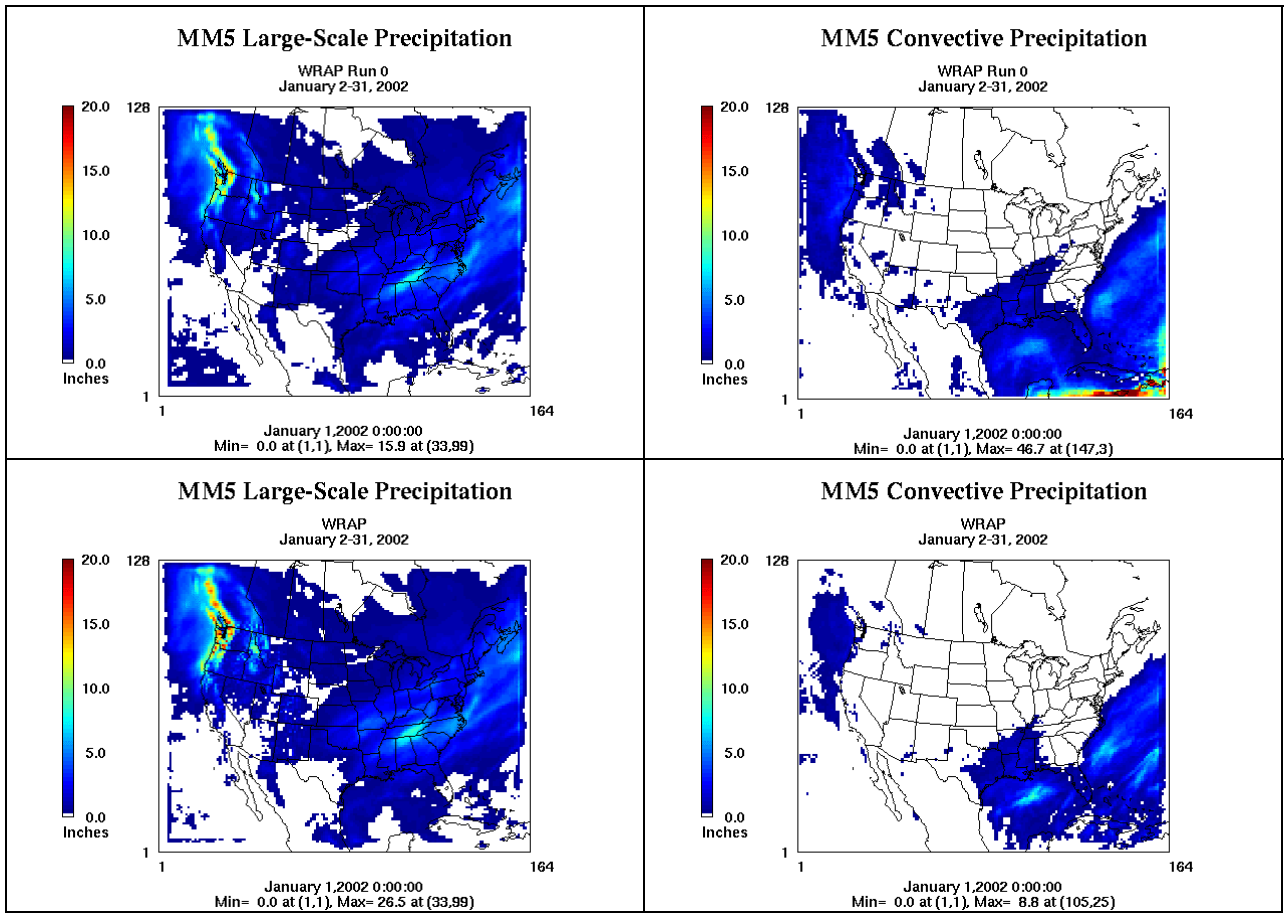

Figure 4-24. Partitioning of January total rainfall into its convective and stratiform components for the WRAP_0 (top) and final WRAP (bottom) 36 km MM5 runs.

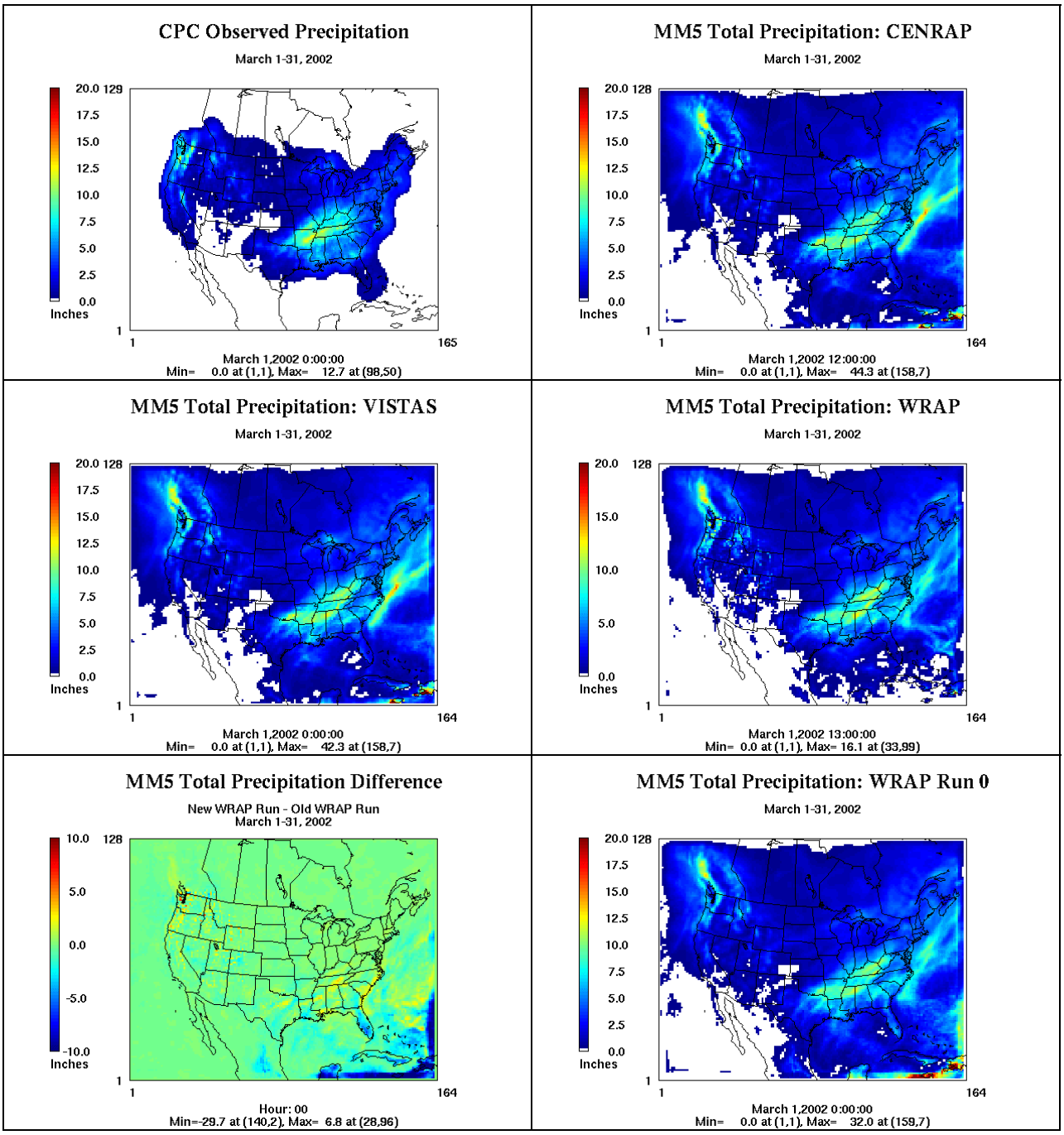

Figure 4-25. March observed (top left) and modeled precipitation for CENRAP (top right), VISTAS (middle left), final WRAP (middle right) and WRAP_0 (bottom right) 2002 36 km simulations and differences between WRAP-0 and final WRAP (bottom left).

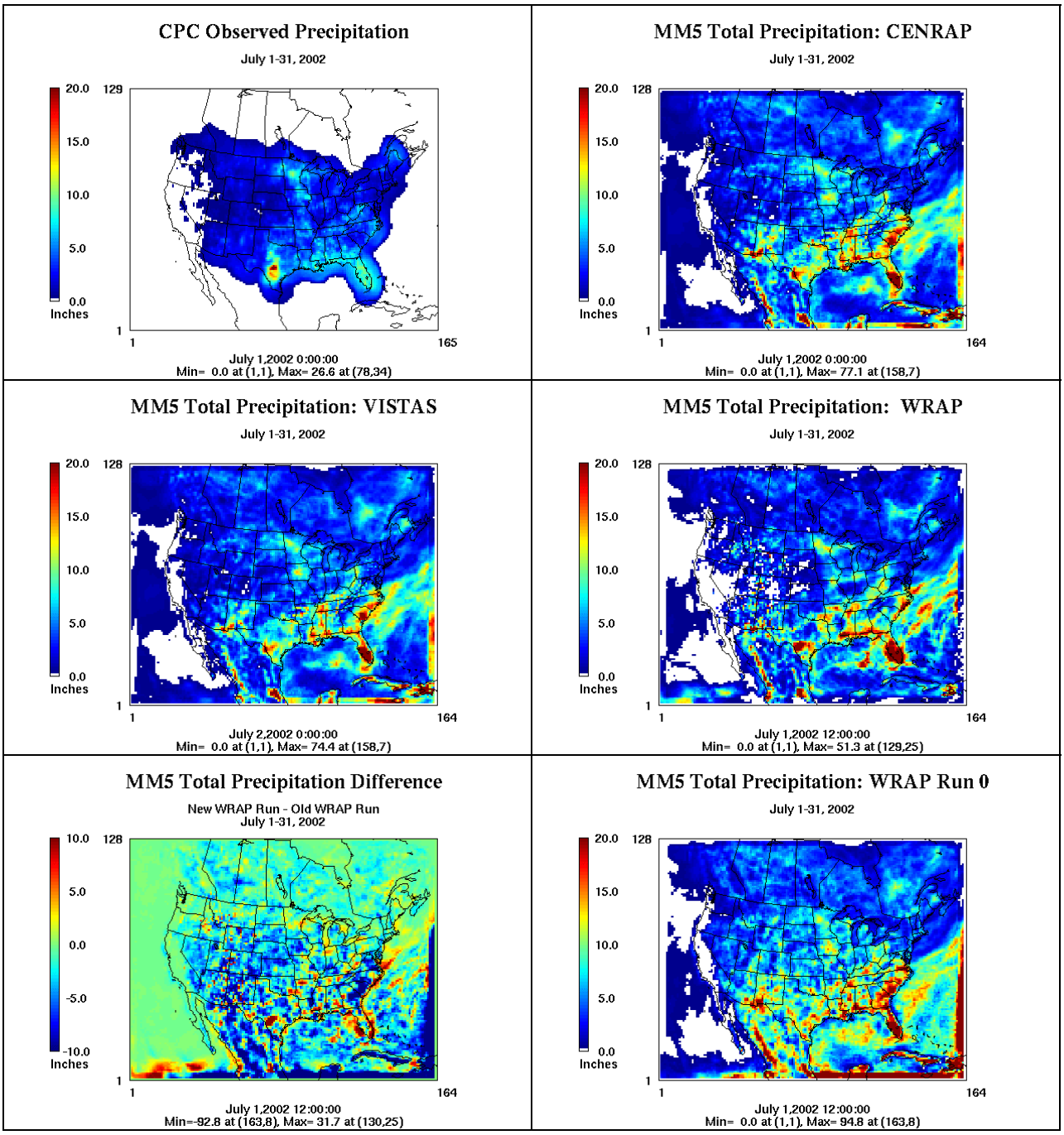

Figure 4-26. July observed (top left) and modeled precipitation for CENRAP (top right), VISTAS (middle left), final WRAP (middle right) and WRAP_0 (bottom right) 2002 36 km simulations and differences between WRAP-0 and final WRAP (bottom left).

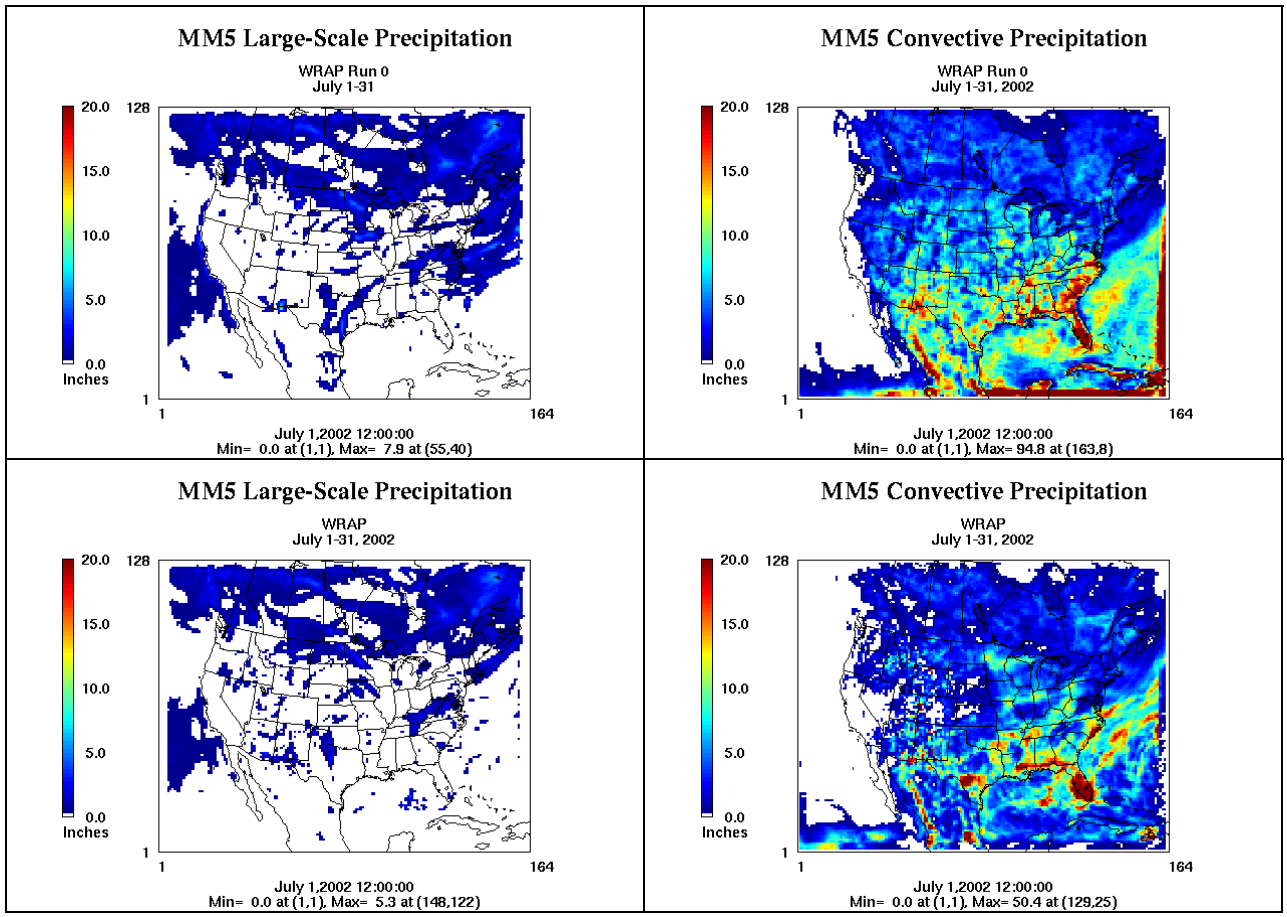

Figure 4-27. Partitioning of July total rainfall into its convective and stratiform components for the WRAP_0 (top) and final WRAP (bottom) 36 km MM5 runs.

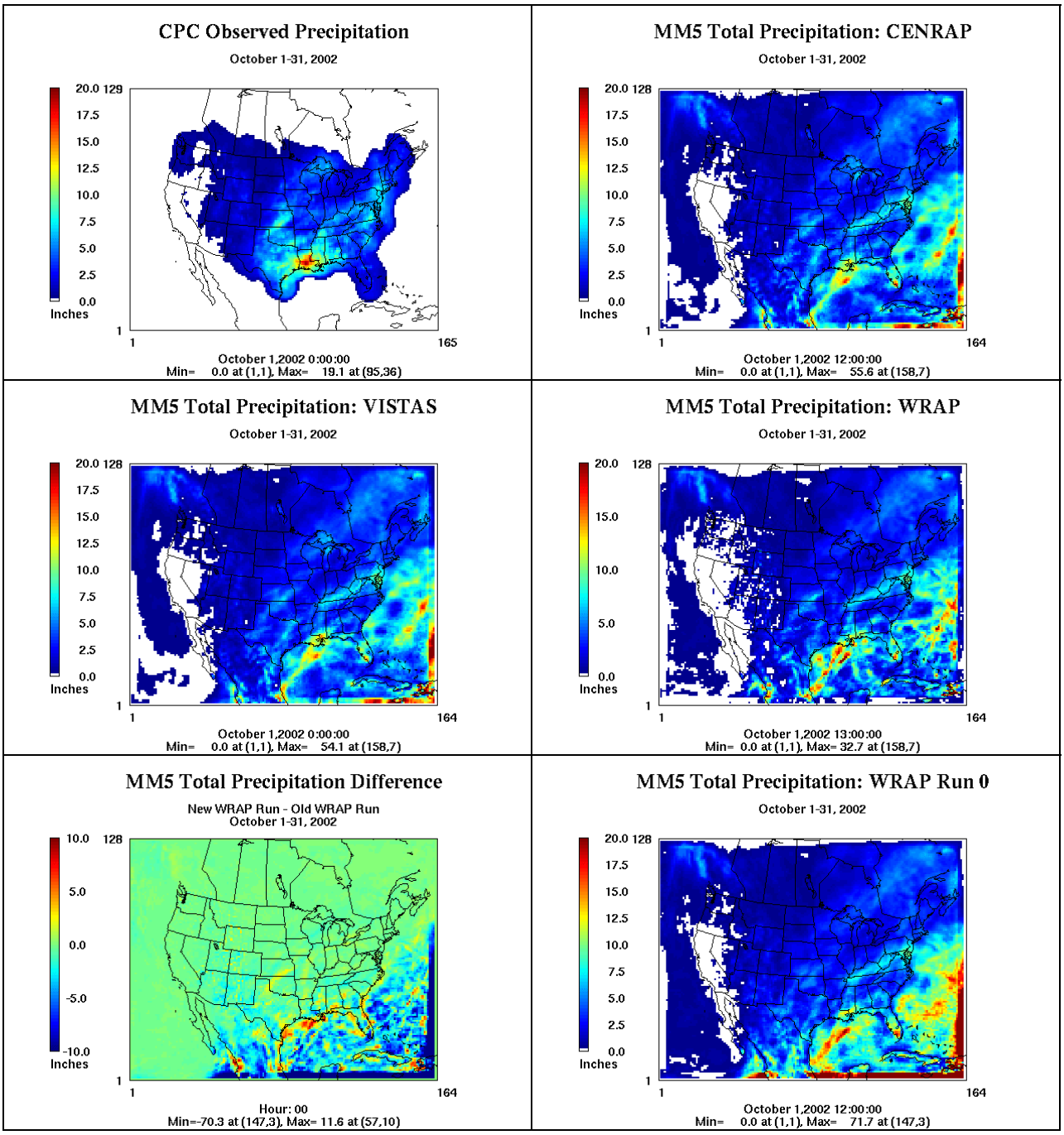

Figure 4-28. October observed (top left) and modeled precipitation for CENRAP (top right), VISTAS (middle left), final WRAP (middle right) and WRAP_0 (bottom right) 2002 36 km simulations and differences between WRAP-0 and final WRAP (bottom left).

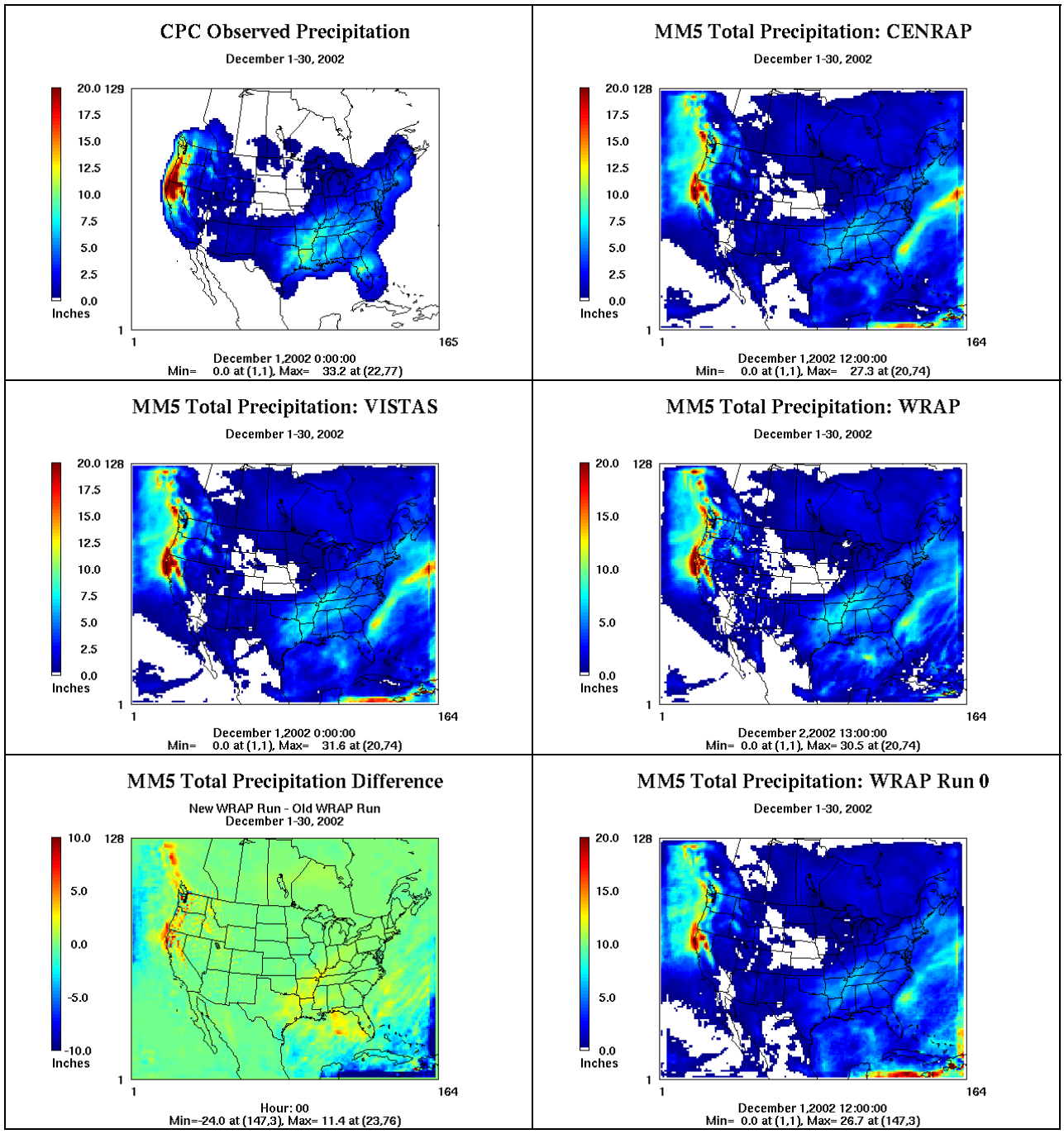

Figure 4-29. December observed (top left) and modeled precipitation for CENRAP (top right), VISTAS (middle left), final WRAP (middle right) and WRAP_0 (bottom right) 2002 36 km simulations and differences between WRAP-0 and final WRAP (bottom left).

UPPER-AIR METEOROLOGICAL EVALUATION

To assess whether MM5 is simulating the vertical structure of the atmosphere with reasonable accuracy in the four 36 km runs, we compared model temperature and dew point soundings with those from a limited number of radiosonde stations. Based on the analysis of surface meteorology and precipitation model performance, it seemed reasonable to focus on the months of January and July. These two months were at extremes of good and poor model performance at the surface.

For the months of January and July, a radiosonde station was selected from each subdomain, and the 00Z and 12Z observed and modeled soundings from each day of the January and July were compared for the CENRAP, VISTAS, WRAP_0, and final WRAP runs. 12Z and 00Z soundings were analyzed in order to examine differences between stable (nocturnal) and unstable (daytime) conditions (00Z and 12Z are 4pm PST/7pm EST and 4am PST/7am PST, respectively). In the following discussion, we will summarize the general features of the upper air soundings found in the extended analysis of January and July. By looking at two months for a limited number of stations, we had a manageable amount of sounding data to analyze, but we caution that it would be unwise to place too much confidence in generalizing these results from two months of data from 11 stations to the entire MM5 domain and the entire year.

Summary of Upper Air Sounding Comparison

In this section, we discuss some common features of the four simulations and some important differences among them. Overall, in all four runs, MM5 simulates the temperature profile more accurately than the dew point temperature profile. Because of its limited resolution, the model tends to have a smoother dew point temperature than the observations, and may miss sudden excursions in the dew point temperature related to the vertical distribution of humidity. In general, MM5 tended to be more accurate in reproducing the 00Z sounding than the 12Z sounding for a given station. This seems to be because MM5 has greater skill in simulating the fully developed afternoon/early evening convective boundary layer than the early morning nocturnal boundary layer, whose physical processes are more complex and less well understood.

In both January and July, the CENRAP and VISTAS runs were better able to simulate PBL temperature inversions than were the two WRAP runs. Figure 4-30 shows the 12Z sounding for Midland, TX (MAF) for January 7, 2002. The CENRAP and VISTAS runs both underestimate the strength of the inversion, but the problem is worse in the WRAP_0 and final WRAP runs. In both WRAP runs, the temperature is colder in the 900-800 mb layer than it is in CENRAP or VISTAS. The dew point temperature sounding in both WRAP runs is also too warm in the 900-825 mb layer; CENRAP and VISTAS also have this problem, but to a lesser extent. Figure 4-31 shows the 12Z sounding for Spokane, WA (OTX) for July 2, 2002. Here, again, MM5 underestimates the inversion strength in all four runs, but the problem is more severe in the two WRAP runs. Note that the WRAP_0 run simulation of the dew point profile is too cold in the 950-750 mb layer, and that CENRAP and VISTAS both do a better job with this profile. The final WRAP run is a significant improvement over WRAP_0 in its handling of the low level dew point profile.

CENRAP and VISTAS dew point temperature profiles were generally closer than that of WRAP_0 to observed profile, and were better able to handle extreme excursions in temperature and dew point profiles in the lower troposphere. The final WRAP run tended to have a more accurate dew point profile than the WRAP_0 run at 12Z, and a temperature profile that was less accurate near the surface, which is reasonable given the lack of surface temperature analysis nudging the final WRAP run. Because of the analysis nudging of surface temperature, the WRAP_0 surface temperature was generally closer than CENRAP or VISTAS to the observed surface temperature, but profiles above surface were less accurate in WRAP_0 than in CENRAP or VISTAS indicating that although the surface temperature nudging improves model performance at the surface it does not for temperatures aloft.

For July soundings with a deep, convecting boundary layer, CENRAP and VISTAS frequently better reproduced the observed temperature inversion at the top of the PBL, while the WRAP_0 run showed a smoother decrease of temperature with increasing altitude. This problem is somewhat ameliorated in the final WRAP run. Figure 4-32 shows the observed and simulated soundings for July 16, 0Z at Midland, TX. The observed sounding shows a deep, dry adiabatic convecting layer with an inversion at 700 mb. Both the CENRAP and VISTAS runs simulate an inversion just above 700 mb, although it is weaker than the observed inversion. The WRAP_0 temperature sounding shows temperature decreasing smoothly with increasing altitude, with no inversion. The final WRAP run still does not manage to produce an inversion, but does a better job of approximating the temperature profile immediately above the observed inversion.

Vertical wind profiles in all four runs were similar, particularly above the boundary layer (this reasonable since all runs were analysis nudged above the PBL). Even in cases where the PBL temperature profiles were different, wind structure was similar. This suggests that local thermal processes are driving the temperature differences, rather than advection.

One issue for all four runs is a discrepancy between the observed and modeled surface pressure. Figure 4-33 shows the 12Z sounding for Flagstaff, AZ (FGZ) for January 4, 2002 for all four runs. The modeled soundings for the CENRAP and VISTAS runs are quite similar. The WRAP_0 sounding is less accurate than either CENRAP or VISTAS, completely missing the inversion near the 700 mb level, and is too warm and too dry in the 800-700 mb layer. The excessive warmth and dryness also occur in the CENRAP and VISTAS runs, but to a lesser degree. The final WRAP run has a more accurate surface pressure than any of the other three runs. This improvement in the surface pressure in the final WRAP run seems to be a general feature of this run, and was noted in both summer and winter soundings. It is unclear what caused this improvement. Like the WRAP_0 run, the final WRAP run does not simulate an inversion near 700 mb, but is closer to the observed temperature and dew point profiles.

In general, the differences between the initial and final WRAP runs varied with season. The two runs differ in their parameterization of cumulus convection and surface observation and analysis nudging. During winter, there was not much difference between the WRAP runs. What differences there were occurred near the surface. The WRAP_0 run tended to have a more accurate temperature profile near the surface, due to its surface analysis nudging. Because cumulus convection is not as active during the winter, the surface temperature differences predominated in winter. There was generally little change in the dew point profile in winter.

In summer, there were larger differences stemming from the difference in parameterization of cumulus convection. Although there was a deterioration in the temperature profile near the surface in the final WRAP run (due to turning off surface temperature nudging), the dew point profile was often more accurate in the final WRAP run (see for example, Figure 4-31)

Also, the use of the Betts-Miller cumulus scheme tended to produce profiles of temperature and dew point indicating that the atmosphere was close to or at saturation above the PBL top inversion. Figure 4-34 shows soundings for Flagstaff, AZ on July 4 and July 6 at 0Z. These soundings are fairly typical for July at this site. The final WRAP run, which uses the Betts-Miller cumulus scheme, is clearly being driven harder toward a saturated profile than the WRAP_0 run. This is likely due to the way the Betts-Miller scheme relaxes toward a particular reference profile as cumulus convection acts to remove grid-scale instability. This type of profile was not observed in the CENRAP or VISTAS runs, which, like WRAP_0, use a Kain-Fritsch scheme.

Overall, the upper air profiles in the CENRAP and VISTAS runs were very similar to one another, and both of these runs had some marked differences with the two WRAP runs (i.e. success at simulating low-level nocturnal inversions or the PBL top inversion). Because the WRAP_0 run used a Kain-Fritsch scheme, as did the CENRAP and VISTAS runs, it seems likely that the differences between the two WRAP runs and the other two runs may be attributed to the different explicit moisture schemes. CENRAP and VISTAS use Reisner I and the WRAP runs used Reisner II. Because simulation of inversions is so critical for visibility modeling, the issue of explicit scheme selection and its relation to the upper air atmospheric structure in MM5 bears further investigation.

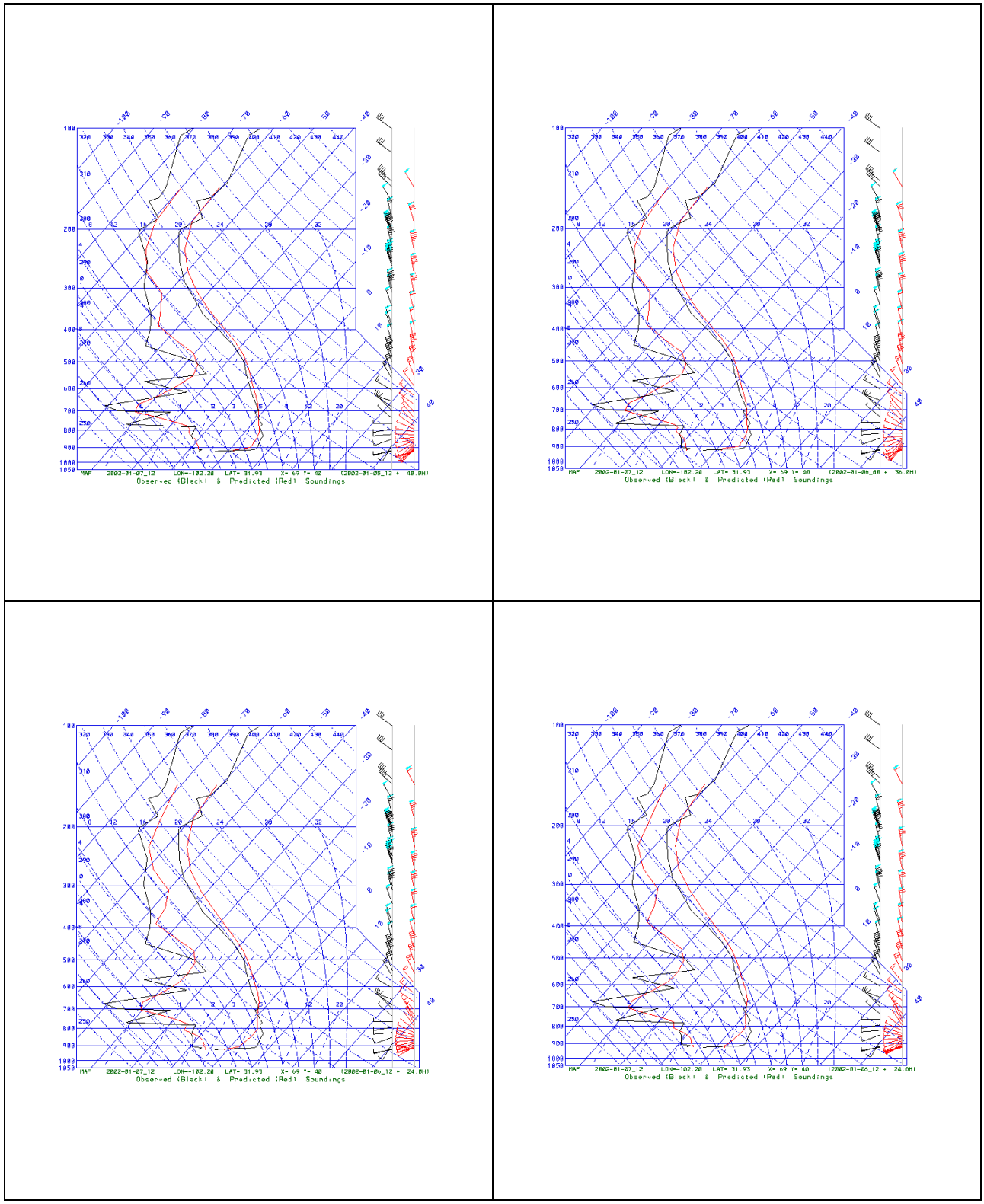

Figure 4-30. January 7, 2002, 12Z Sounding for Midland, TX. Upper left hand panel: CENRAP, Upper right hand panel: VISTAS, Lower left hand panel: WRAP_0, Lower right hand panel: Final WRAP.

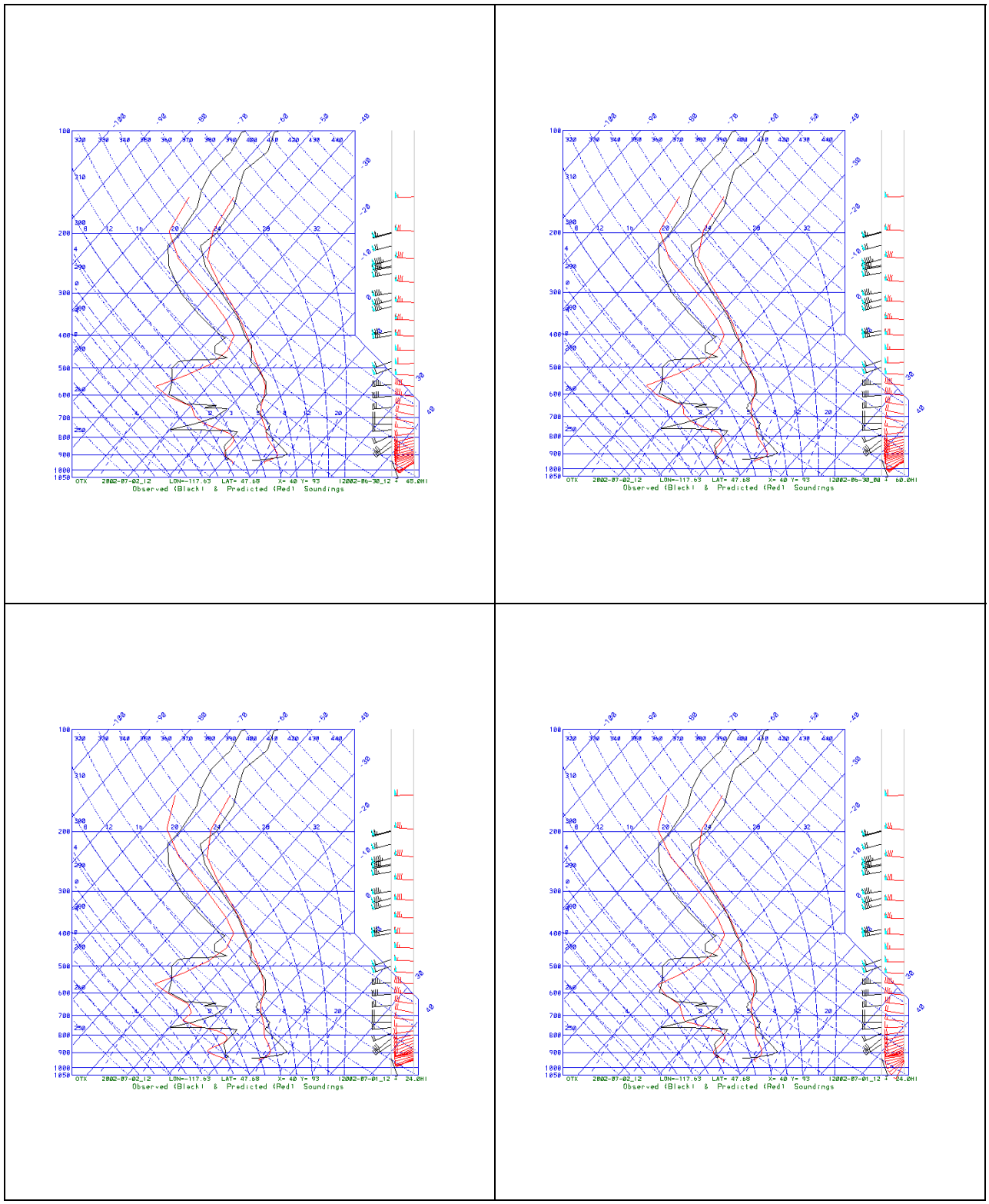

Figure 4-31. July 2, 2002, 12Z Sounding for Spokane, WA. Upper left hand panel: CENRAP, Upper right hand panel: VISTAS, Lower left hand panel: WRAP_0, Lower right hand panel: Final WRAP.

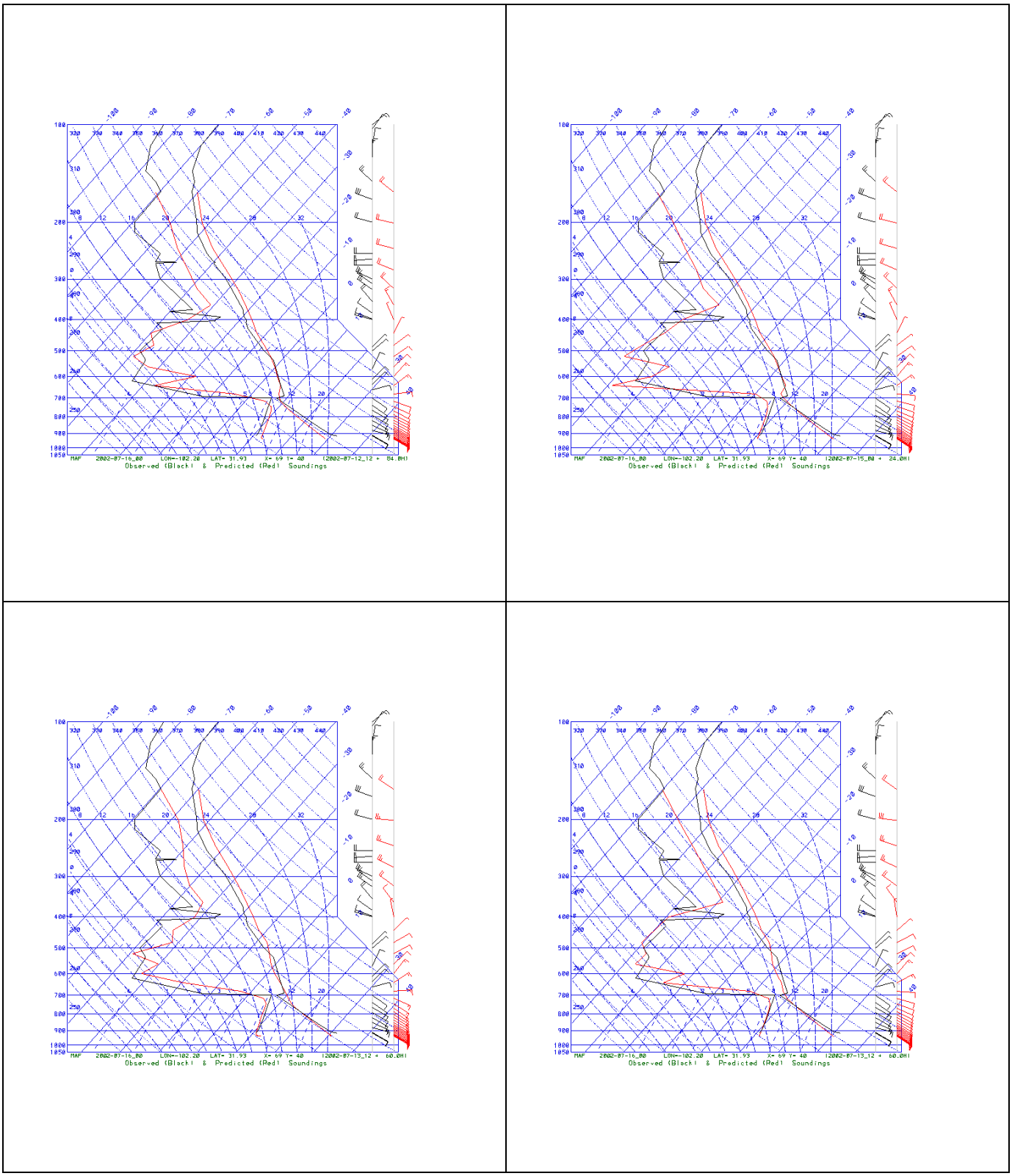

Figure 4-32. July 16, 2002, 0Z Sounding for Midland, TX. Upper left hand panel: CENRAP, Upper right hand panel: VISTAS, Lower left hand panel: WRAP_0, Lower right hand panel: Final WRAP.

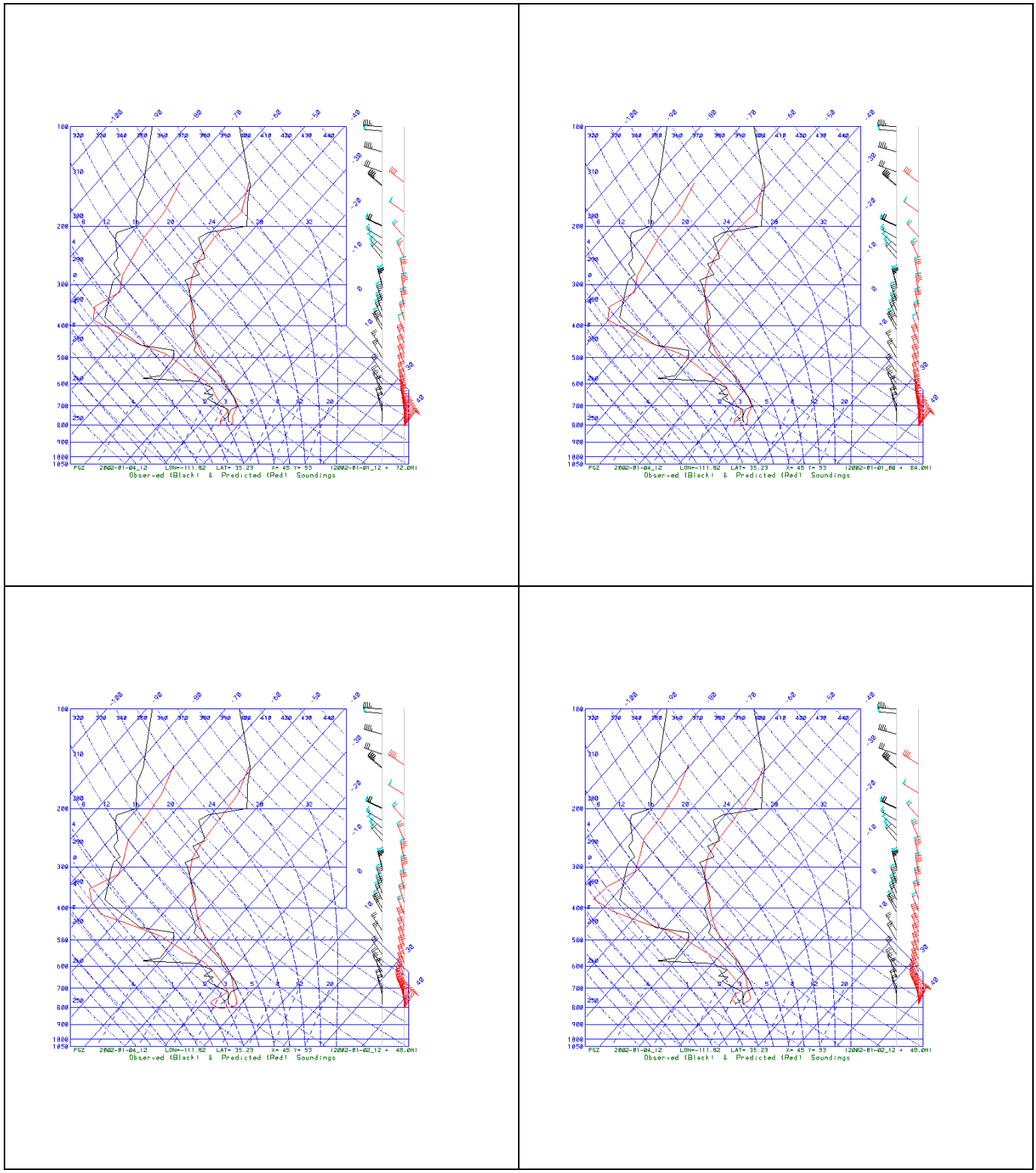

Figure 4-33. January 4, 2002, 12Z Sounding for Flagstaff, AZ. Upper left hand panel: CENRAP, Upper right hand panel: VISTAS, Lower left hand panel: WRAP_0, Lower right hand panel: Final WRAP.

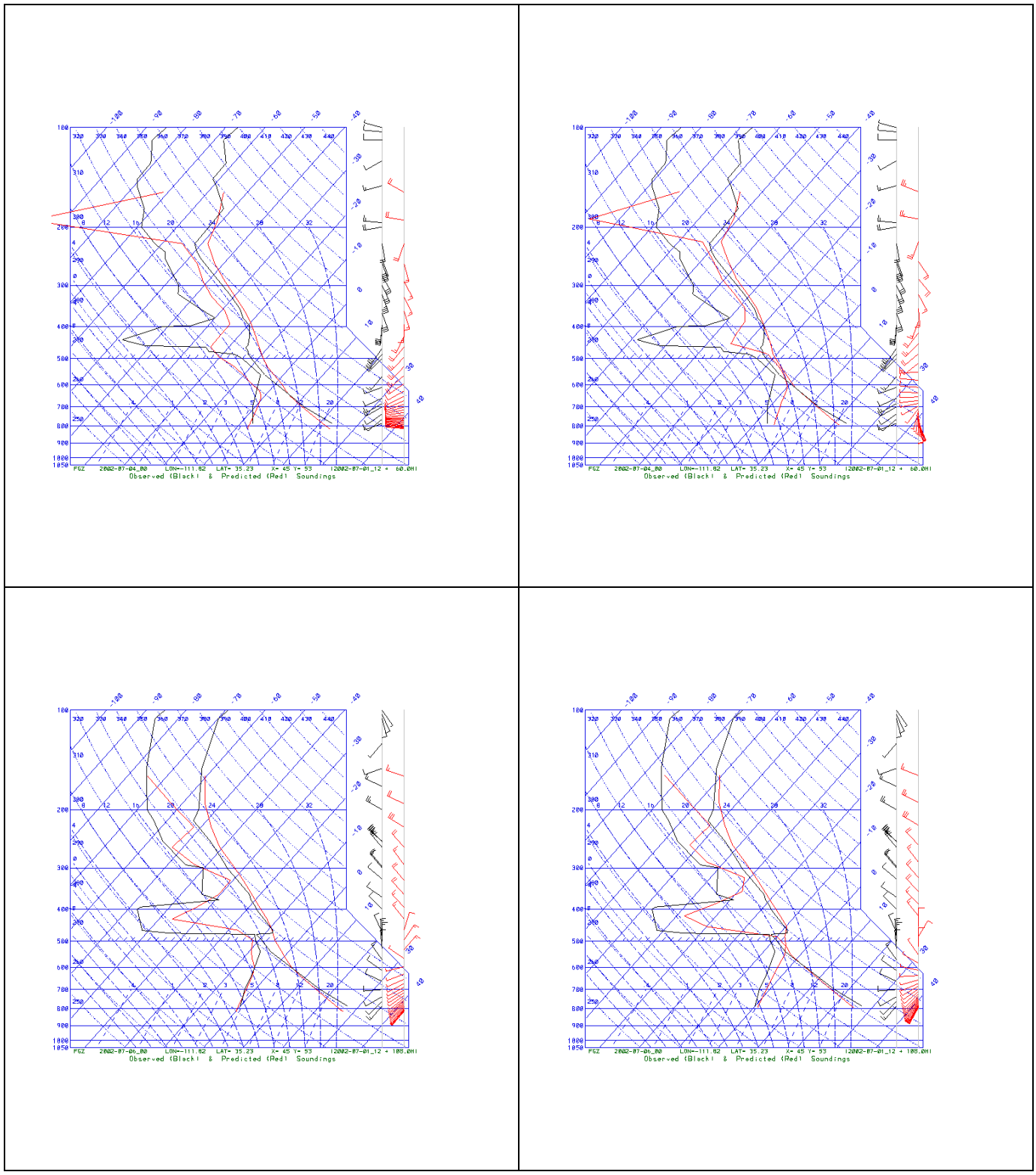

Figure 4-34. July 4,6, 2002, 0Z Soundings for Flagstaff, AZ. Upper left hand panel: WRAP_0 July 4, Upper right hand panel: Final WRAP July 4, Lower left hand panel: WRAP_0 July 6, Lower right hand panel: Final WRAP Jul 6.

5.0 EVALUATION OF THE 2002 12 KM WRAP MM5 SIMULATION

In Section 5, we assess the performance of the 2002 12 km MM5 run. The evaluation of the MM5 surface meteorological variables using METSTAT is presented first. This is followed by a comparison of the CPC observed and MM5 predicted precipitation patterns.

As in the analysis of the 36 km run in Section 4, soccerplots were generated and evaluated for each month of the year 2002 for the 12 km grid. On these soccerplots, we also include for reference the data from the 36 km run, so that we can see the effects on performance of running at higher resolution and with different cumulus parameterization (none, in the case of the 12 km run). As in the 36 km run, MM5's performance was qualitatively different during summer and winter, with the spring and fall seasons serving as transitional periods between winter and summer. January and July were found to be representative of MM5's performance during the winter and summer seasons. For brevity, we include only January and July in this report. The main strengths and weaknesses of the 2002 12 km MM5 run are captured in the January and July plots. For both months, we show a soccerplot for the subdomains in the western U.S. (the 12 km WRAP region). The subdomain definitions are the same as those used in the 36 km analysis (shown in Figure 3-1).

Surface MM5 Meteorological Performance

Figure 5-1 displays the January soccerplot for temperature for the 12 km and 36 km final WRAP runs. For both 36 km and 12 km runs, all subdomains fall outside the benchmark. For the North, SW, and desertSW subdomains, the 12 k run offers a small improvement in performance, but not enough to move the subdomains significantly closer to the benchmark.

The wind performance for January is shown in Figure 5-2. For all four subdomains, the 12 km wind performance falls within the benchmark and is a slight improvement over the 36 km performance. For humidity (Figure 5-3), both the 36 km and 12 km runs are within the benchmark, with the 12 km run again showing a slight improvement over the 36 km run.

Figure 5-4 shows the July soccerplot for temperature for the 12 km and 36 km final WRAP runs. The overall pattern, with the PacificNW, North, and desertSW having a cold bias and SW having a warm bias is the same as in January. The cold bias of the PacificNW, North, and desertSW subdomains has decreased relative to January, while the warm bias of the SW subdomain has increased. The bias for the PacificNW, North, and desertSW subdomains is now within the benchmark. Although the temperature error is not within the benchmark for these three subdomains, it has been reduced relative to January. This is a significant improvement in performance relative to the WRAP_0 36 km and 12 km runs, and is a result of the final WRAP run configuration's optimization for improvement of the July humidity and temperature performance. The reduction in cold bias in these three subdomains is very likely a result of the improved precipitation field relative to the WRAP_0 run. The final WRAP run does not have the same overprediction of rainfall, so that surface energy is no longer spuriously diverted into evaporating the excess moisture from the surface, and is more properly used in heating the surface. This reduces the cold bias in the surface temperature. The July 12 km run is less skillful overall than the 36 km run for temperature. It is unclear why this is so.

In Figure 5-5, we show the wind performance soccerplot for July. The wind performance is slightly degraded relative to January, but is relatively close to the benchmark, and is improved relative to the 12 km WRAP_0 run (not shown). The 12 km run shows an improvement in wind speed RMSE, but not wind direction error relative to the 36 km run.

For July humidity (Figure 5-6), all subdomains fall within the benchmark. This is in marked contrast to the WRAP_0 run (not shown), where strong wet biases in the occurred North and desertSW subdomains due to excessive convective rainfall in those regions. The large wet biases of these two subdomains, which placed them well outside the range of results for previous meteorological databases used in air quality applications, was a major motivation for performing the final WRAP run with a different configuration than the WRAP_0 run. Relative to January, the 12 km results for July humidity show that the humidity error has increased for all four subdomains with the onset of summer, and that the wet bias has increased for the PacificNW, desertSW, and North subdomains. This is consistent with the overestimate of July rainfall (true even in the final WRAP run; see Figures 5-7E and 5-7F) in these three subdomains. The SW subdomain does not receive much rainfall in July, and its humidity bias does not change much relative to January, although its humidity error increases.

In summary, the METSTAT surface analysis shows that the 12 km run is within or near performance benchmarks for wind and humidity over the annual cycle of 2002, but the temperature results fall outside the bias benchmark. It is possible that this is due to terrain resolution effects. The final 12 km WRAP run has been significantly improved in terms of its surface performance relative to the WRAP_0 12 km run.

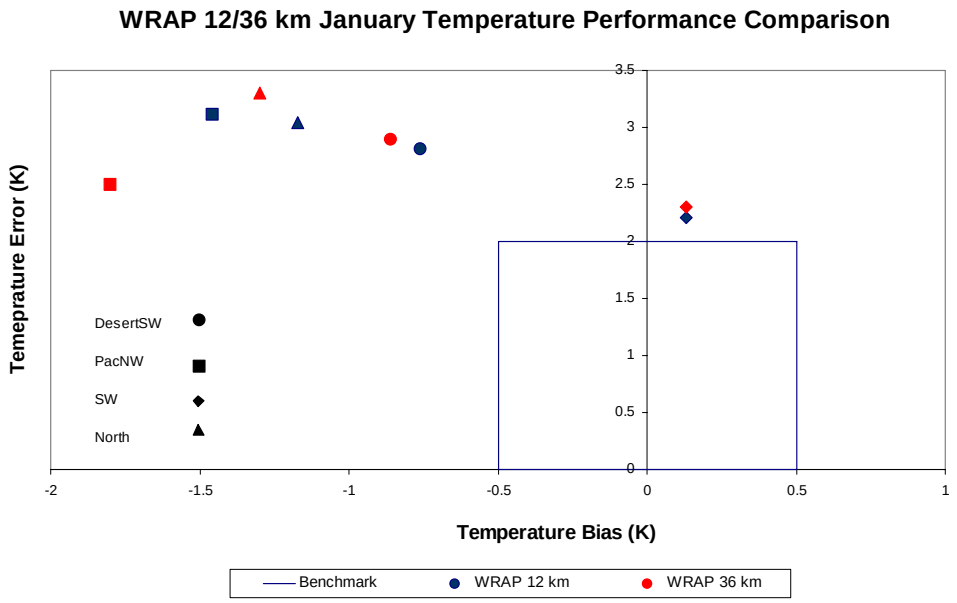

Figure 5-1. 36km/12 km surface temperature comparison soccerplot for January.

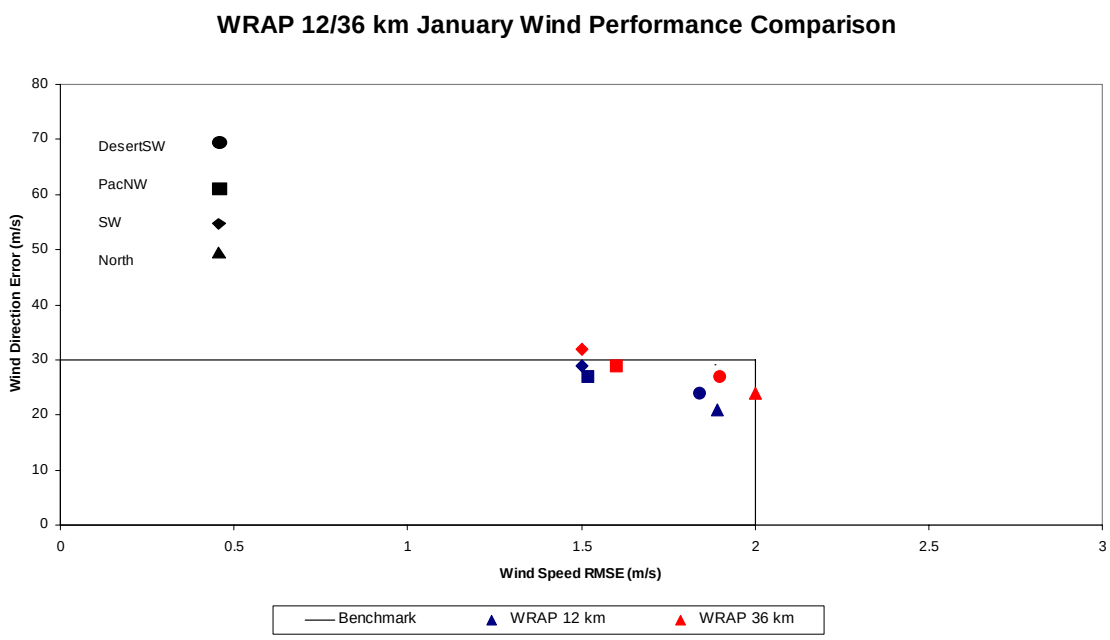

Figure 5-2. 36km/12 km surface wind comparison soccerplot for January.

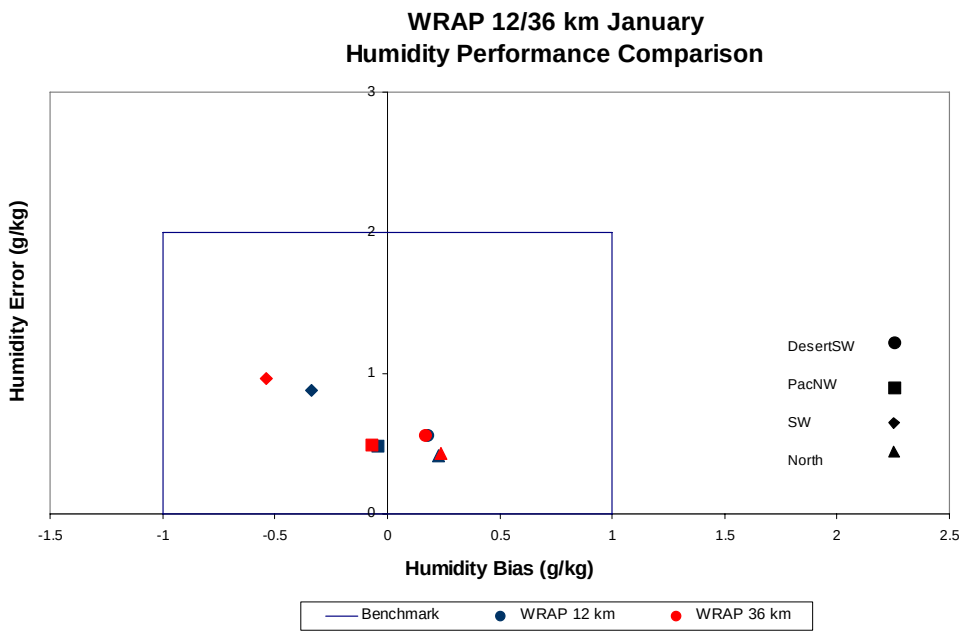

Figure 5-3. 36km/12 km surface humidity comparison soccerplot for January.

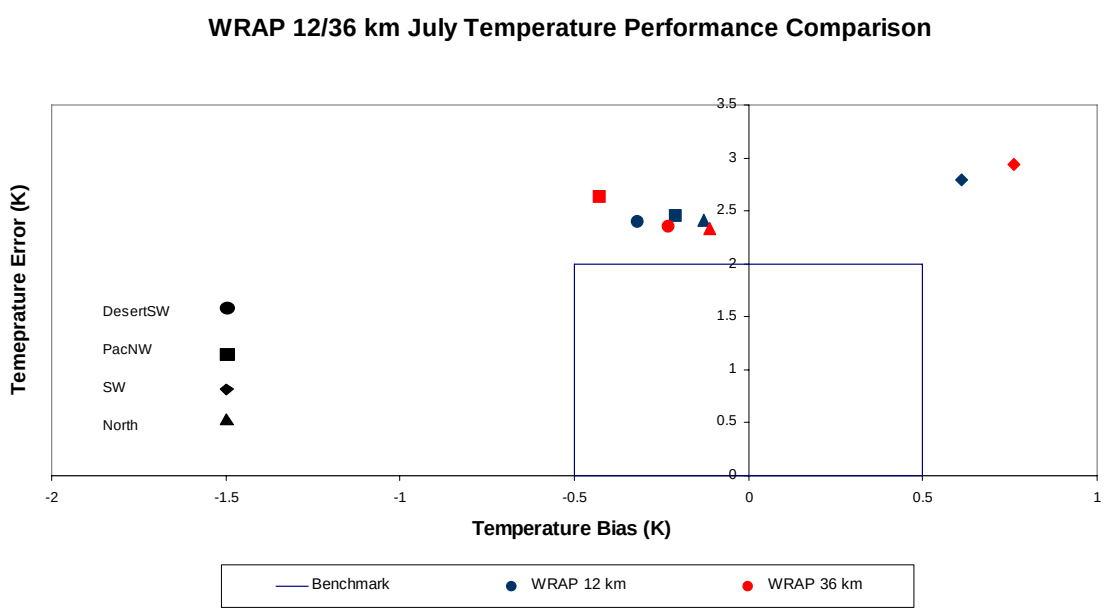

Figure 5-4. 36km/12 km surface temperature comparison soccerplot for July.

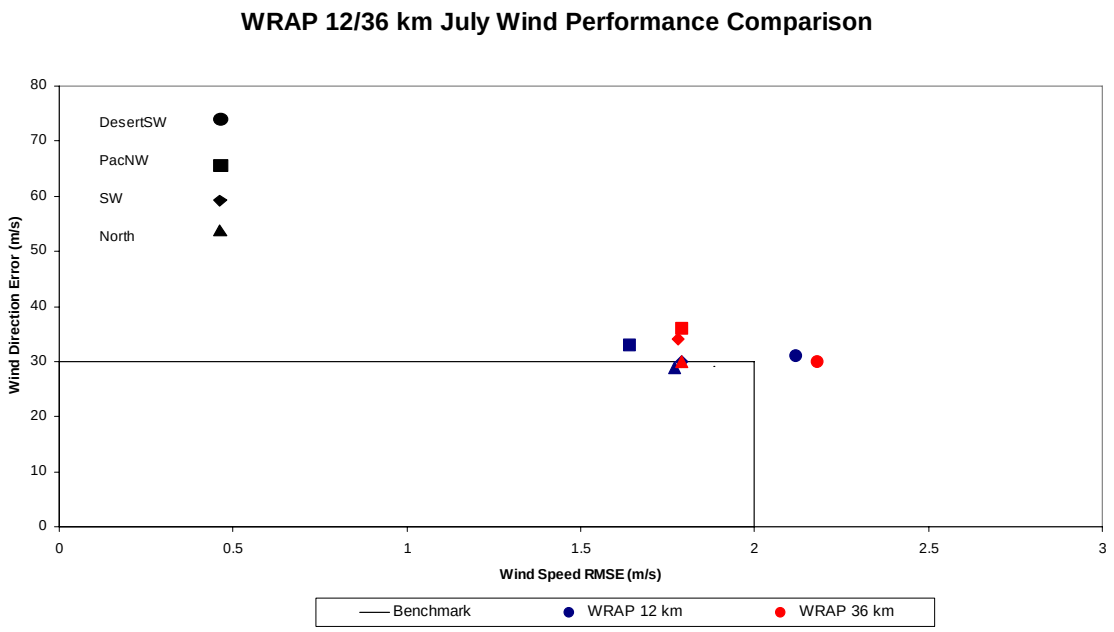

Figure 5-5. 36km/12 km surface wind comparison soccerplot for July.

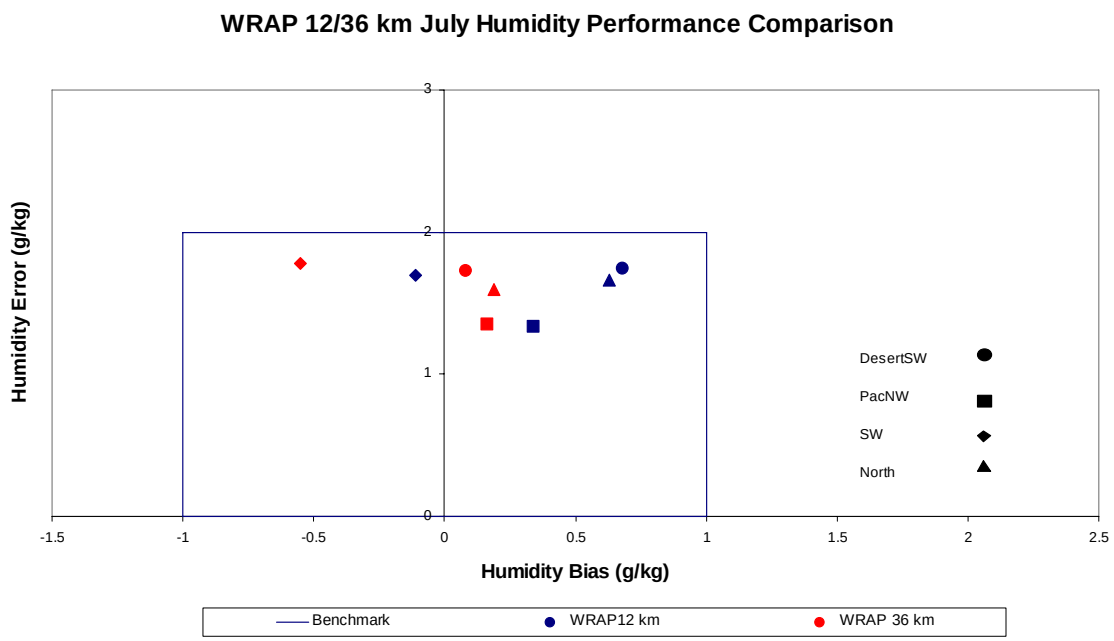

Figure 5-6. 36km/12 km humidity wind comparison soccerplot for July.

EVALUATION OF PRECIPITATION IN THE 2002 12 KM WRAP RUN

In this section, we evaluate the precipitation performance of the 2002 12 km MM5 run. The original WRAP_0 run had a strong positive precipitation bias over the WRAP region on both the 36 km and 12 km grids. In tandem with the 36 km sensitivity tests described in Kemball-Cook et al. (2004b), a second series of sensitivity tests was performed on the 12 km grid to identify a more optimal configuration for that domain. The best choice for cumulus scheme turned out to be no cumulus parameterization, as discussed in Section 2. Figures 5-7A through 5-7J show a comparison of the observed CPC precipitation and the MM5-predicted precipitation over the course of the final WRAP 2002 12 km simulation.

In January (Figures 5-7A and 5-7B), the agreement between the overall predicted and observed precipitation pattern is reasonably good. MM5 picks up the precipitation maxima over the mountain ranges in the Pacific Northwest, although rainfall amounts are too high over both ranges. There is excessive precipitation in the cenrapN and North regions, and MM5 under predicts the rainfall over the central California coast. In March (Figure 5-7C and 5-7D), the model again over predicts the rainfall over the mountain ranges of the Pacific Northwest. Aside from this, however, MM5 shows impressive skill.

As we move from winter to summer and convective rainfall becomes more important, MM5's forecast skill deteriorates. Figure 5-7E and Figure 5-7F show the observed and modeled precipitation for July. Although MM5 does a reasonable job with the SW and PacNW subdomains, where little or no rain falls, there is a general overprediction of rainfall. This is consistent with the wet bias seen in the surface humidity soccerplot for July (Figure 5-6) for all subdomains except SW. In general, the model does a good job with the overall precipitation pattern, but individual maxima are over predicted. The model is running with no convective parameterization. This means that in order for convection to occur, the entire 12 km grid column must saturate and be unstable. In the real world, convective updrafts tend to be smaller than 12 km across. This may mean that it is relatively difficult for modeled convection to be initiated, and it becomes unrealistically intense when it does occur because an unphysical amount of instability has been allowed to build up in the convecting grid cell.

As fall arrives, and the partitioning of rainfall moves toward an increase in the stratiform component, MM5's performance improves. Figures 5-7G and 5-7H show the observed and modeled precipitation for October. In October, the model underestimates precipitation in the banded features over the cenrapS region, but otherwise agrees reasonably well with observations.

MM5's December performance (Figures 5-7I and 5-7J) is similar to that of January. The model simulates the overall pattern of precipitation over the Pacific Northwest, but overestimates the intensity of the maxima. As in January, precipitation over the North and cenrapN subdomains is overestimated. Otherwise, the model does a good job of simulating the December precipitation field.

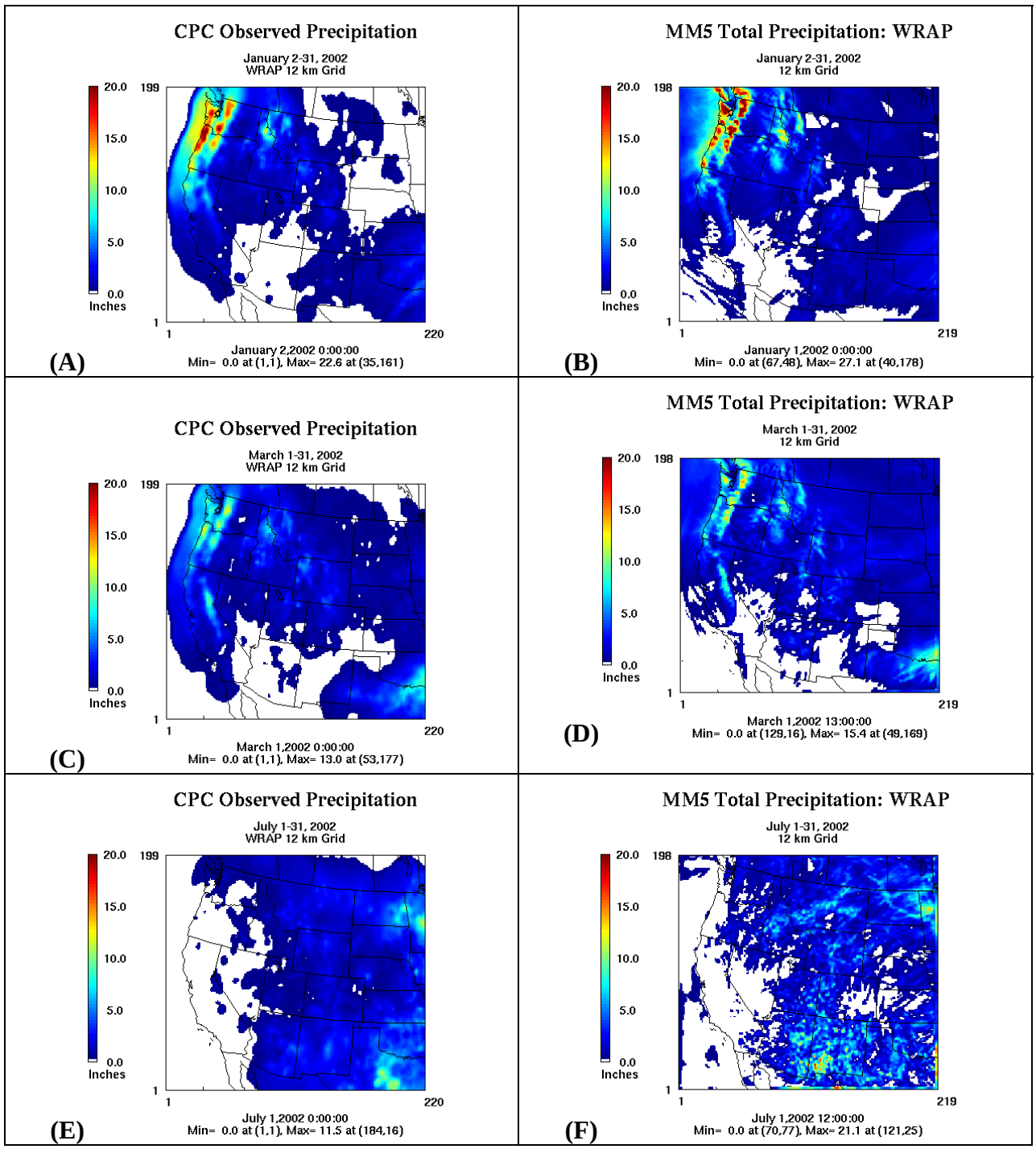

Figure 5-7. Annual cycle in 12 km MM5 precipitation.
(A) January CPC Observed Precipitation.
(B) January MM5 Predicted Total Precipitation.
(C) March CPC Observed Precipitation.
(D) March MM5 Predicted Total Precipitation.
(E) July CPC Observed Precipitation.
(F) July MM5 Predicted Total Precipitation.

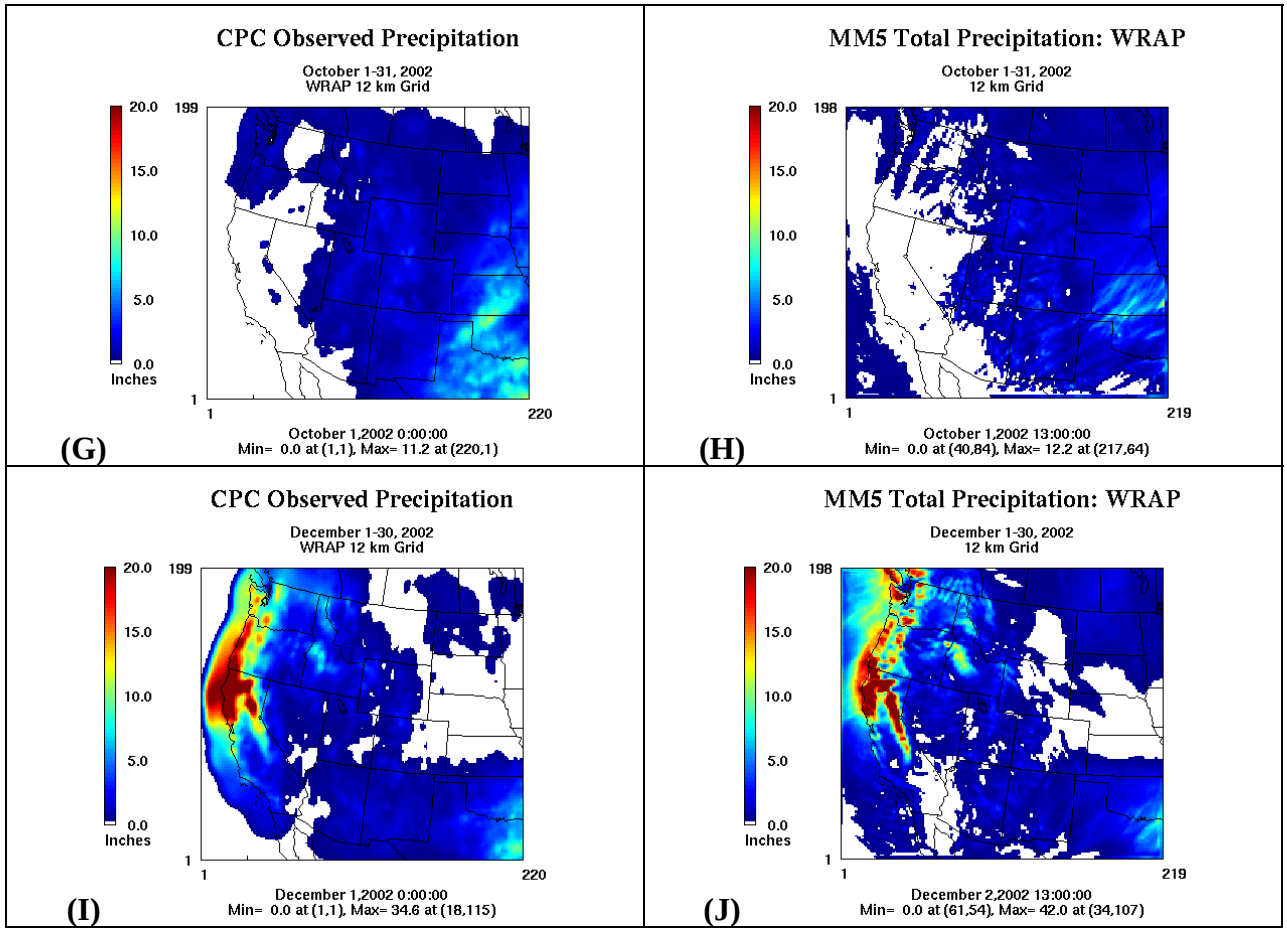

Figure 5-7. Annual cycle in 12 km MM5 precipitation. (Cont.)
(G) October CPC Observed Precipitation
(H) October MM5 Predicted Total Precipitation.
(I) December CPC Observed Precipitation.
(J) December MM5 Predicted Total Precipitation.

In summary, MM5 predicts the precipitation on the 12 km grid with reasonable skill over most of the annual cycle. The performance is better in winter than in summer. Throughout the year, the model tends to overpredict precipitation maxima, but does a good job in simulating the overall precipitation pattern. The final WRAP run exhibits better skill than the WRAP_0 run, particularly in July. The modeled rainfall is still excessive in July, but the severity of the over prediction and the corresponding biases in July surface temperature and humidity have lessened.

6.0 CONCLUSIONS

At both 12 km and 36 km resolutions, the final WRAP 2002 MM5 simulation produced results that are generally within the range of meteorological model results that have been used in the past for air quality applications. The final 36 km and 12 km runs represent a significant improvement in performance over the original WRAP_0 run.

In the 36 km WRAP_0 simulation, surface humidity and temperature fields fell within the benchmarks for much of the year for most subdomains, but the model had a marked cold, wet bias especially during the summer in the western subdomains. This was unfortunate since the 20% worst visibility days at western class I areas that are of high interest to WRAP generally occur during the summer. The surface wind field failed to meet the performance benchmark for the entire year in subdomains PacificNW, SW, North, and DesertSW. Given that the purpose of this run was to provide a database for a 12 km nested MM5 run and air quality modeling studies centered on the western U.S., this was a serious problem.

The final 36 km and 12 km MM5 runs show improvement in the modeled precipitation fields, particularly in the summer in the southwest. With a reduction in the overprediction of summer convective rainfall in both 36 and 12 km runs, wet biases in the surface humidity and cold biases in the surface temperature are now smaller in the North and DesertSW subdomains in the summer months. Although the final WRAP run was optimized for summer performance over the WRAP region, the winter performance in the west did not deteriorate significantly. There was a small loss of accuracy, particularly in temperature and humidity in the east, but this region is not the focus of WRAP. To summarize, the final WRAP 36 km run:

- Saw its surface wind performance improve significantly throughout the year due to observational nudging of surface winds

- Showed significant improvement in summer rainfall and surface humidity performance in the WRAP region

- Did worse than the original WRAP run for humidity performance in the east. We can attribute both of these changes in performance to reduction in the areal coverage in rainfall and increase in convective rainfall in active cells (and therefore surface moisture) that we get in going from the KF scheme to the Betts-Miller scheme.

- Showed improved temperature performance in summer in the west, and slightly worse performance in winter

- Showed a small overall degradation in performance in the east. Some of this was the results of eliminating surface analysis nudging of temperature and moisture that was done in WRAP_0

Our comparison of the two 36 km WRAP runs (WRAP_0 and final WRAP) to the CENRAP and VISTAS 36 km continental-scale annual runs showed:

- Overall, VISTAS performed best in the simulation of surface temperature and humidity. It had no outliers and has no serious performance problems at any particular time or subdomain. It is unclear whether this is due to the explicit moist physics or convection schemes or to the interaction between them, or to effects of differences in FDDA.

- The final WRAP run performed best for surface winds throughout the year.

- For precipitation, the four runs were similar in terms of performance if the whole 36 km domain and the entire year are considered. Over the WRAP region, however, the final WRAP run performs best, with the smallest overprediction of convective rain of all runs. The overprediction of convective precipitation was most severe in the WRAP_0 run.

- For upper air structure, CENRAP and VISTAS performed best and were similar to one another. We attribute this to the use of the Reisner I scheme, as the simulation of, for example, the PBL inversion was relatively insensitive to the change of convection scheme in the WRAP runs.

- In all four runs, MM5 performed better in winter than in summer.

- Although WRAP_0 had the best surface statistics for temperature, its upper air performance was worst of all four runs, which suggests that analysis nudging of surface temperature and humidity is counterproductive. The final WRAP run, which did not have analysis nudging of surface temperature and humidity, had larger errors in its temperature structure in the lowest levels of the atmosphere, but had a more realistic dew point profile and a smaller surface pressure bias than any of the four runs.

- Based on the upper air soundings, the most serious problem is the difficulty MM5 has is establishing the observed PBL structure. MM5 has trouble getting the PBL depth right, particularly in the stable nocturnal case. Also, the model's difficulty in simulating the observed fine structure of the dew point temperature profile and the overall level of saturation in the lower troposphere is cause for concern. It is important that model produce cloud decks at the correct height. Errors in humidity and cloud prediction will have a negative impact on the accuracy of downwelling solar radiation, cause errors in the temperature profile and the surface fluxes, affect chemistry, and make it difficult for the PM model to perform properly.

We conclude, based on the results of this study, that the final 36 km and 12 km WRAP MM5 runs exhibit reasonably good performance and are certainly within the bounds of other meteorological databases used for prior air quality modeling efforts. It is therefore reasonable to proceed with their use as inputs for visibility modeling.

REFERENCES

Arakawa, A. and J.M. Chen. 1987. "Closure assumptions in the cumulus parameterization problem." Short and Medium Range Numerical Prediction, Collection of Papers. Presented at the WMO/IUGG NWP Symposium, Tokyo. T. Matsuno, Ed. Meteor. Soc. Japan. 107-131.

ATMET. 2003. "MM5 Simulations for TexAQS 2000 Episode, Task 3: Sensitivities to modifications of the MRF PBL scheme." Prepared for the Houston Advanced Research Center, The Woodlands, TX, and the Texas Commission on Environmental Quality, Austin, TX, by ATMET, LLC, Boulder, CO September.

Betts, A. and M. Miller. 1986. A new convective adjustment scheme. Part I: Observational and theoretical basis. *Quart. J. Roy. Meteor. Soc.* **99**, 178-196.

Emery, C.A. and E. Tai. 2001. "Enhanced meteorological modeling and performance evaluation for two Texas ozone episodes." Prepared for the Texas Natural Resource Conservation Commission, by ENVIRON International Corporation.

Emery, C.A., S. Kemball-Cook, Y. Jia, Z. Wang and R. Morris. 2004. "2002 MM5 Model Evaluation: 12 vs 36 km Results." Presented at May 24-25, 2004 National RPO Modeling Meeting, Denver, CO. http://www.cleanairinfo.com/rpomodelingdenver.

ENVIRON and UCR. 2004. "2002 Annual MM5 Simulations to Support WRAP CMAQ Visibility Modeling for the Section 308 SIP/TIP". Draft Protocol. ENVIRON International Corporation and the University of California at Riverside. April. Available at: http://pah.cert.ucr.edu/aqm/308/mm5.shtml.

Gochis, D., W.J. Shuttleworth, and Z.L. Yang. 2002. "Sensitivity of the modeled North American Monsoon regional climate to convective parameterization." *Mon. Wea. Rev.*, **130**, 1282-1298.

Hart, J. 2000. "Forecasting studies using hourly model-generated data." Final Report to the COMET Outreach Program.

Kemball-Cook, S., Y. Jia, C. Emery, R. Morris, Z. Wang and G. Tonnesen. 2004a. "2002 Annual MM5 36 km simulation to support WRAP CMAQ Visibility Modeling for the Section 308 SIP/TIP: Preliminary Report on the Initial 2002 36 km MM5 Simulation Performed during 2003." ENVIRON International Corporation and UC Riverside. Available at: http://pah.cert.ucr.edu/aqm/308/mm5.shtml. March.

Kemball-Cook, S., Y. Jia, C. Emery, R. Morris, Z. Wang and G. Tonnesen. 2004. "MM5 Configuration Sensitivity Testing" ENVIRON International Corporation and UC Riverside. Available at: http://pah.cert.ucr.edu/aqm/308/mm5.shtml. December.

Kemball-Cook, S., Y. Jia, C. Emery, R. Morris. 2004. "Response to Comments on Sensitivity Report" Available at: http://pah.cert.ucr.edu/aqm/308/mm5.shtml. December.

Molinari, J. and M. Dudek. 1992. "Parameterization of convective precipitation in mesoscale numerical models." *Mon. Wea. Rev*., **120**, 326-344.

Molinari, J.. 1993. "An overview of cumulus parameterization in mesoscale models. The Representation of Cumulus Convection in Numerical Models. K. Emanuel and D. Raymond, Ed. *American Meteorological Society*, p.155-159.

Morris, R.E., S. Kemball-Cook, Y. Jia, C. Emery and Z. Wan. 2004a. "2002 MM5 36 km Evaluation." Presented at WRAP Regional Modeling Center Workshop, Tempe Arizona." Available at: http://pah.cert.ucr.edu/aqm/308/meetings/Jan_2004_ppt/040128Modeling_Workshop_Agenda.htm. January.

Morris, R.E., B. Koo, S. Lau, T.W. Tesche, D. McNally, C. Loomis, G. Stella, G. Tonnesen and Z. Wang. 2004b. "VISTAS Emissions and Air Quality Modeling – Task 4cd Report: Model Performance Evaluation and Model Sensitivity Tests for Three Phase I Episodes. ENVIRON International Corporation, Novato, California. September. Available at: http://pah.cert.ucr.edu/vistas/docs.shtml .

Olerud, D. and A. Sims. 2003. "MM5 Sensitivity modeling in support of VISTAS (Visibility Improvement-State and Tribal Association)." Prepared for Mike Abraczinskas of the VISTAS Technical Analysis Workgroup by Baron Advanced Meteorological Systems, LLC.

Pleim, J.E. and J.S. Chang. 1992: A non-local closure model for vertical mixing in the convective boundary layer. *Atmos. Env*., **26A**, pp. 965-981.

Randall, D.A., M. Khairoutdinov, A. Arakawa, and W. Grabowski. 2003. Breaking the cloud parameterization deadlock. *Bull. Am. Met Soc.* **84**, 1547-1564.

Reisner, J.R., R.M. Rasmussen, and R.T. Bruintjes. 1998. Explicit forcing of supercooled liquid water in winter storms using the MM5 mesoscale model. Quart. J. Roy. Met. Soc., 124B, pp. 1071-1107.

Tesche, T.W. et al. 2002. "Operational Evaluation of the MM5 Meteorological Model over the Continental United States: Protocol for Annual and Episodic Evaluation." Prepared for the U.S. Environmental Protection Agency, Office of Air Quality Planning and Standards, prepared by Alpine Geophysics, LLC, Ft. Wright, KY.

Xiu, A. and J.E. Pleim. 2000. Development of a land surface model. Part I: application in a mesoscale meteorology model. *J. App. Met.*, **40**, pp. 192-209.